CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION

Related Applications

This application claims priority to prior filed U.S. Provisional Patent Application Serial No. 60/141031, filed June 25, 1999, U.S. Provisional Patent Application Serial No. 60/143694, filed July 14, 2000, and U.S. Provisional Patent Application Serial No. 60/151778, filed August 31, 1999. This application also claims priority to German Application No. 19931418.7, filed July 8, 1999, German Application No. 19932124.8, filed July 9, 1999, German Application No. 19932126.4, filed July 9, 1999, German Application No. 19932127.2, filed July 9, 1999, German Application No. 19932133.7, filed July 9, 1999, German Application No. 19932207.4, filed July 9, 1999, German Application No. 19932208.2, filed July 9, 1999, German Application No. 19932225.2, filed July 9, 1999, German Application No. 19932229.5, filed July 9, 1999, German Application No. 19932914.1, filed July 9, 1999, German Application No. 19933006.9, filed July 9, 1999, German Application No. 19940765.7, filed August 27, 1999, German Application No. 19940768.1, filed August 27, 1999, German Application No. 19940831.9, filed August 27, 1999, German Application No. 19940832.7, filed August 27, 1999, German Application No. 19941385.1, filed August 31, 1999, German Application No. 19941396.7, filed August 31, 1999, and German Application No. 19942087.4, filed September 3, 1999. The entire contents of all of the aforementioned applications are hereby expressly incorporated herein by this reference.

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as homeostasis and adaptation (HA) proteins.

C. glutamicum is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The HA nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, *e.g.*, by fermentation processes. Modulation of the expression of the HA nucleic acids of the invention, or modification of the sequence of the HA nucleic acid molecules of the invention, can be used to modulate the production of one or more fine chemicals from a microorganism (*e.g.*, to improve the yield or production of one or more fine chemicals from a Corynebacterium or Brevibacterium species).

The HA nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

The HA nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms. Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered Corynebacterium or Brevibacterium species.

The HA proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or in the ability of this microorganism to adapt to

different environmental conditions. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g.*, *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (*e.g.* phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (*e.g.*, amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (*e.g.*, regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as HA proteins, which are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or of

participating in the ability of this microorganism to adapt to different environmental conditions. Nucleic acid molecules encoding an HA protein are referred to herein as HA nucleic acid molecules. In a preferred embodiment, an HA protein participates in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or possesses a *C. glutamicum* enzymatic or proteolytic activity. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an HA protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of HA-encoding nucleic acids (*e.g.*, DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, *e.g.*, sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an HA activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (*e.g.*, an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire

amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (*e.g.*, an HA fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (*i.e.*, the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* HA protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, *e.g.*, recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an HA protein by culturing the host cell in a suitable medium. The HA protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an HA gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated HA sequence as a transgene. In another embodiment, an endogenous HA gene within the genome of the microorganism has been altered, *e.g.*, functionally disrupted, by homologous recombination with an altered HA gene. In another embodiment, an endogenous or introduced HA gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for

the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

In another aspect, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

Still another aspect of the invention pertains to an isolated HA protein or a portion, *e.g.*, a biologically active portion, thereof. In a preferred embodiment, the isolated HA protein or portion thereof can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions. In another preferred embodiment, the isolated HA protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions.

The invention also provides an isolated preparation of an HA protein. In preferred embodiments, the HA protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated HA protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated HA protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of HA proteins also have one or more of the HA bioactivities described herein.

The HA polypeptide, or a biologically active portion thereof, can be operatively linked to a non-HA polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the HA protein alone. In other preferred embodiments, this fusion protein participates in the maintenance of homeostasis in *C. glutamicum*, or performs a function involved in the adaptation of this microorganism to different environmental conditions. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

In another aspect, the invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an HA nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an HA nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates HA protein activity or HA nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* processes involved in cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activities. The agent which modulates HA protein activity can be an agent which stimulates HA protein activity or HA nucleic acid expression. Examples of agents which stimulate HA protein activity or HA nucleic acid expression include small molecules, active HA proteins, and nucleic acids encoding HA proteins that have been introduced into the cell. Examples of agents which inhibit HA activity or expression include small molecules and antisense HA nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant HA gene into a cell, either maintained on a separate plasmid or integrated into the genome of

the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

Detailed Description of the Invention

The present invention provides HA nucleic acid and protein molecules which are involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (*e.g.*, where overexpression or optimization of activity of a protein involved in the production of a fine chemical (*e.g.*, an enzyme) has a direct impact on the yield, production, and/or efficiency of production of a fine chemical from the modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (*e.g.*, where modulation of the activity or number of copies of a *C. glutamicum* aromatic or aliphatic modification or degradation protein results in an increase in the viability of *C. glutamicum* cells, which in turn permits increased production in a large-scale culture setting). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described *e.g.* in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (*e.g.*, arachidonic acid), diols (*e.g.*, propane diol, and butane diol), carbohydrates (*e.g.*, hyaluronic acid and trehalose), aromatic compounds (*e.g.*, aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of

Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane *et al.* (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine,

valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm *et al.* (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them.

Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (*e.g.*, polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate

(GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (*e.g.*, pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system. The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and

biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (*i.e.*, AMP) or as coenzymes (*i.e.*, FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (*e.g.* Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (*e.g.*, thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (*e.g.*, ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (*e.g.*, IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto *et al.*, (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Maintenance of Homeostasis in *C. glutamicum* and Environmental Adaptation

The metabolic and other biochemical processes by which cells function are sensitive to environmental conditions such as temperature, pressure, solute concentration, and availability of oxygen. When one or more such environmental condition is perturbed or altered in a fashion that is incompatible with the normal functioning of these cellular processes, the cell must act to maintain an intracellular

environment which will permit them to occur despite the hostile extracellular environment. Gram positive bacterial cells, such as *C. glutamicum* cells, have a number of mechanisms by which internal homeostasis may be maintained despite unfavorable extracellular conditions. These include a cell wall, proteins which are able to degrade possibly toxic aromatic and aliphatic compounds, mechanisms of proteolysis whereby misfolded or misregulated proteins may be rapidly destroyed, and catalysts which permit intracellular reactions to occur which would not normally take place under the conditions optimal for bacterial growth.

Aside from merely surviving in a hostile environment, bacterial cells (*e.g. C. glutamicum* cells) are also frequently able to adapt such that they are able to take advantage of such conditions. For example, cells in an environment lacking desired carbon sources may be able to adapt to growth on a less-suitable carbon source. Also, cells may be able to utilize less desirable inorganic compounds when the commonly utilized ones are unavailable. *C. glutamicum* cells possess a number of genes which permit them to adapt to utilize inorganic and organic molecules which they would normally not encounter under optimal growth conditions as nutrients and precursors for metabolism. Aspects of cellular processes involved in homeostasis and adaptation are further explicated below.

A. Modification and Degradation of Aromatic and Aliphatic Compounds

Bacterial cells are routinely exposed to a variety of aromatic and aliphatic compounds in nature. Aromatic compounds are organic molecules having a cyclic ring structure, while aliphatic compounds are organic molecules having open chain structures rather than ring structures. Such compounds may arise as by-products of industrial processes (*e.g.*, benzene or toluene), but may also be produced by certain microorganisms (*e.g.*, alcohols). Many of these compounds are toxic to cells, particularly the aromatic compounds, which are highly reactive due to the high-energy ring structure. Thus, certain bacteria have developed mechanisms by which they are able to modify or degrade these compounds such that they are no longer hazardous to the cell. Cells may possess enzymes that are able to, for example, hydroxylate, isomerize, or methylate aromatic or aliphatic compounds such that they are either rendered less toxic, or such that the modified form is able to be processed by standard cellular waste and degradation pathways. Also, cells may possess enzymes which are able to specifically degrade one or more such potentially hazardous substance, thereby protecting the cell. Principles and examples of these types of modification and degradation processes in bacteria are described in several publications, *e.g.*, Sahm, H. (1999) "Procaryotes in Industrial Production" in Lengeler, J.W. *et al.*, eds.

Biology of the Procaryotes, Thieme Verlag: Stuttgart; and Schlegel, H.G. (1992) Allgemeine Mikrobiologie, Thieme: Stuttgart).

Aside from simply inactivating hazardous aromatic or aliphatic compounds, many bacteria have evolved to be able to utilize these compounds as carbon sources for continued metabolism when the preferred carbon sources of the cell are not available. For example, Pseudomonas strains able to utilize toluene, benzene, and 1,10-dichlorodecane as carbon sources are known (Chang, B.V. *et al.* (1997) *Chemosphere* 35(12): 2807-2815; Wischnak, C. *et al.* (1998) *Appl. Environ. Microbiol.* 64(9): 3507-3511; Churchill, S.A. *et al.* (1999) *Appl. Environ. Microbiol.* 65(2): 549-552). There are similar examples from many other bacterial species which are known in the art.

The ability of certain bacteria to modify or degrade aromatic and aliphatic compounds has begun to be exploited. Petroleum is a complex mixture of chemicals which includes aliphatic molecules and aromatic compounds. By applying bacteria having the ability to degrade or modify these toxic compounds to an oil spill, for example, it is possible to eliminate much of the environmental damage with high efficiency and low cost (see, for example, Smith, M.R. (1990) "The biodegradation of aromatic hydrocarbons by bacteria" *Biodegradation* 1(2-3): 191-206; and Suyama, T. *et al.* (1998) "Bacterial isolates degrading aliphatic polycarbonates," *FEMS Microbiol. Lett.* 161(2): 255-261).

B. Metabolism of Inorganic Compounds

Cells (*e.g.*, bacterial cells) contain large quantities of different molecules, such as water, inorganic ions, and organic substances (*e.g.*, proteins, sugars, and other macromolecules). The bulk of the mass of a typical cell consists of only 4 types of atoms: carbon, oxygen, hydrogen, and nitrogen. Although they represent a smaller percentage of the content of a cell, inorganic substances are equally as important to the proper functioning of the cell. Such molecules include phosphorous, sulfur, calcium, magnesium, iron, zinc, manganese, copper, molybdenum, tungsten, and cobalt. Many of these compounds are critical for the construction of important molecules, such as nucleotides (phosphorous) and amino acids (nitrogen and sulfur). Others of these inorganic ions serve as cofactors for enzymic reactions or contribute to osmotic pressure. All such molecules must be taken up by the bacterium from the surrounding environment.

For each of these inorganic compounds it is desirable for the bacterium to take up the form which can be most readily used by the standard metabolic machinery of the cell. However, the bacterium may encounter environments in which these preferred forms are not readily available. In order to survive under these circumstances, it is important for bacteria to have additional biochemical mechanisms which are able to convert less metabolically active but readily available forms of these inorganic compounds to ones which may be used

in cellular metabolism. Bacteria frequently possess a number of genes encoding enzymes for this purpose, which are not expressed unless the desired inorganic species are not available. Thus, these genes for the metabolism of various inorganic compounds serve as another tool which bacteria may use to adapt to suboptimal environmental conditions.

After carbon, the most important element in the cell is nitrogen. A typical bacterial cell contains between 12-15% nitrogen. It is a constituent of amino acids and nucleotides, as well as many other important molecules in the cell. Further, nitrogen may serve as a substitute for oxygen as a terminal electron acceptor in energy metabolism. Good sources of nitrogen include many organic and inorganic compounds, such ammonia gas or ammonia salts (*e.g.*, NH_4Cl, $(NH_4)_2SO_4$, or NH_4OH), nitrates, urea, amino acids, or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract, etc. Ammonia nitrogen is fixed by the action of particular enzymes: glutamate dehydrogenase, glutamine synthase, and glutamine-2-oxoglutarate aminotransferase. The transfer of amino-nitrogen from one organic molecule to another is accomplished by the aminotransferases, a class of enzymes which transfer one amino group from an alpha-amino acid to an alpha-keto acid. Nitrate may be reduced via nitrate reductase, nitrite reductase, and further redox enzymes until it is converted to molecular nitrogen or ammonia, which may be readily utilized by the cell in standard metabolic pathways.

Phosphorous is typically found intracellularly in both organic and inorganic forms, and may be taken up by the cell in either of these forms as well, though most microorganisms preferentially take up inorganic phosphate. The conversion of organic phosphate to a form which the cell can utilize requires the action of phosphatases (*e.g.*, phytases, which hydrolyze phyate-yielding phosphate and inositol derivatives). Phosphate is a key element in the synthesis of nucleic acids, and also has a significant role in cellular energy metabolism (*e.g.*, in the synthesis of ATP, ADP, and AMP).

Sulfur is a requirement for the synthesis of amino acids (*e.g.*, methionine and cysteine), vitamins (*e.g.*, thiamine, biotin, and lipoic acid) and iron sulfur proteins. Bacteria obtain sulfur primarily from inorganic sulfate, though thiosulfate, sulfite, and sulfide are also commonly utilized. Under conditions where these compounds may not be readily available, many bacteria express genes which enable them to utilize sulfonate compounds such as 2-aminosulfonate (taurine) (Kertesz, M.A. (1993) "Proteins induced by sulfate limitation in *Escherichia coli*, *Pseudomonas putida*, or *Staphylococcus aureus*." *J. Bacteriol.* 175: 1187-1190).

Other inorganic atoms, *e.g.*, metal or calcium ions, are also critical for the viability of cells. Iron, for example, plays a key role in redox reactions and is a cofactor of iron-sulfur proteins, heme proteins, and cytochromes. The uptake of iron into bacterial cells may be accomplished by the action of siderophores, chelating agents

which bind extracellular iron ions and translocate them to the interior of the cell. For reference on the metabolism of iron and other inorganic compounds, see: Lengeler *et al.* (1999) Biology of Prokaryotes, Thieme Verlag: Stuttgart; Neidhardt, F.C. *et al.*, eds. *Escherichia coli* and *Salmonella*. ASM Press: Washington, D.C.; Sonenshein, A.L. *et al.*, eds. (199?) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM Press: Washington, D.C.; Voet, D. and Voet, J.G. (1992) Biochemie, VCH: Weinheim; Brock, T.D. and Madigan, M.T. (1991) Biology of Microorgansisms, 6th ed. Prentice Hall: Englewood Cliffs, p. 267-269; Rhodes, P.M. and Stanbury, P.F. Applied Microbial Physiology – A Practical Approach, Oxford Univ. Press: Oxford.

C. Enzymes and Proteolysis

The intracellular conditions for which bacteria such as *C. glutamicum* are optimized are frequently not conditions under which many biochemical reactions would normally take place. In order to make such reactions proceed under physiological conditions, cells utilize enzymes. Enzymes are proteinaceous biological catalysts, spatially orienting reacting molecules or providing a specialized environment such that the energy barrier to a biochemical reaction is lowered. Different enzymes catalyze different reactions, and each enzyme may be the subject of transcriptional, translational, or posttranslational regulation such that the reaction will only take place under appropriate conditions and at specified times. Enzymes may contribute to the degradation (*e.g.*, the proteases), synthesis (*e.g.*, the synthases), or modification (*e.g.*, transferases or isomerases) of compounds, all of which enable the production of necessary compounds within the cell. This, in turn, contributes to the maintenance of cellular homeostasis.

However, the fact that enzymes are optimized for activity under the physiological conditions at which the bacterium is most viable means that when environmental conditions are perturbed, there is a significant possibility that enzyme activity will also be perturbed. For example, changes in temperature may result in aberrantly folded proteins, and the same is true for changes of pH – protein folding is largely dependent on electrostatic and hydrophobic interactions of amino acids within the polypeptide chain, so any alteration to the charges on individual amino acids (as might be brought about by a change in cellular pH) may have a profound effect on the ability of the protein to correctly fold. Changes in temperature effectively change the amount of kinetic energy that the polypeptide molecule possesses, which affects the ability of the polypeptide to settle into a correctly folded, energetically stable configuration. Misfolded proteins may be harmful to the cell for two reasons. First, the aberrantly folded protein may have a similarly aberrant activity, or no activity

whatsoever. Second, misfolded proteins may lack the conformational regions necessary for proper regulation by other cellular systems and thus may continue to be active but in an uncontrolled fashion.

The cell has a mechanism by which misfolded enzymes and regulatory proteins may be rapidly destroyed before any damage occurs to the cell: proteolysis. Proteins such as those of the la/lon family and those of the Clp family specifically recognize and degrade misfolded proteins (see, *e.g.*, Sherman, M.Y., Goldberg, A.L. (1999) *EXS* 77: 57-78 and references therein and Porankiewicz J. (1999) *Molec. Microbiol.* 32(3): 449-58, and references therein; Neidhardt, F.C., *et al.* (1996) *E. coli* and *Salmonella*, ASM Press: Washington, D.C. and references therein; and Pritchard, G.G., and Coolbear, T. (1993) *FEMS Microbiol.* Rev. 12(1-3): 179-206 and references therein). These enzymes bind to misfolded or unfolded proteins and degrade them in an ATP-dependent manner. Proteolysis thus serves as an important mechanism employed by the cell to prevent damage to normal cellular functions upon environmental changes, and it further permits cells to survive under conditions and in environments which would otherwise be toxic due to misregulated and/or aberrant enzyme or regulatory activity.

Proteolysis also has important functions in the cell under optimal environmental conditions. Within normal metabolic processes, proteases aid in the hydrolysis of peptide bonds, in the catabolism of complex molecules to provide necessary degradation products, and in protein modification. Secreted proteases play an important role in the catabolism of external nutrients even prior to the entry of these compounds into the cell. Further, proteolytic activity itself may serve regulatory functions; sporulation in *B. subtilis* and cell cycle progression in *Caulobacter* spp. are known to be regulated by key proteolytic events in each of these species (Gottesman, S. (1999) *Curr. Opin. Microbiol.* 2(2): 142-147). Thus, proteolytic processes are key for cellular survival under both suboptimal and optimal environmental conditions, and contribute to the overall maintenance of homeostasis in cells.

D. Cell Wall Production and Rearrangements

While the biochemical machinery of the cell may be able to readily adapt to different and possibly unfavorable environments, cells still require a general mechanism by which they may be protected from the environment. For many bacteria, the cell wall affords such protection, and also plays roles in adhesion, cell growth and division, and transport of desired solutes and waste materials.

In order to function, cells require intracellular concentrations of metabolites and other molecules that are substantially higher than those of the surrounding media. Since these metabolites are largely prevented from leaving the cell due to the presence of the

hydrophobic membrane, the tendency of the system is for water molecules to enter the cell from the external medium such that the interior concentrations of solutes match the exterior concentrations. Water molecules are readily able to cross the cellular membrane, and this membrane is not able to withstand the resulting swelling and pressure, which may lead to osmotic lysis of the cell. The rigidity of the cell wall greatly improves the ability of the cell to tolerate these pressures, and offers a further barrier to the unwanted diffusion of these metabolites and desired solutes from the cell. Similarly, the cell wall also serves to prevent unwanted material from entering the cell.

The cell wall also participates in a number of other cellular processes, such as adhesion and cell growth and division. Due to the fact that the cell wall completely surrounds the cell, any interaction of the cell with its surroundings must be mediated by the cell wall. Thus, the cell wall must participate in any adherence of the cell to other cells and to desired surfaces. Further, the cell cannot grow or divide without concomitant changes in the cell wall. Since the protection that the wall affords requires its presence during growth, morphogenesis and multiplication, one of the key steps in cell division is cell wall synthesis within the cell such that a new cell divides from the old. Thus, frequently cell wall biosynthesis is regulated in tandem with cell growth and cell division (see, *e.g.*, Sonenshein, A.L. et al, eds. (1993) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM: Washington, D.C.).

The structure of the cell wall varies between gram-positive and gram-negative bacteria. However, in both types, the fundamental structural unit of the wall remains similar: an overlapping lattice of two polysaccharides, N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM) which are cross-linked by amino acids (most commonly L-alanine, D-glutamate, diaminopimelic acid, and D-alanine), termed 'peptidoglycan'. The processes involved in the synthesis of the cell wall are known (see, *e.g.*, Michal, G., ed. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York).

In gram-negative bacteria, the inner cellular membrane is coated by a single-layered peptidoglycan (approximately 10 nm thick), termed the murein-sacculus. This peptidoglycan structure is very rigid, and its structure determines the shape of the organism. The outer surface of the murein-sacculus is covered with an outer membrane, containing porins and other membrane proteins, phospholipids, and lipopolysaccharides. To maintain a tight association with the outer membrane, the gram-negative cell wall also has interspersed lipid molecules which serve to anchor it to the surrounding membrane.

In gram-positive bacteria, such as *Corynebacterium glutamicum*, the cytoplasmic membrane is covered by a multi-layered peptidoglycan, which ranges from 20-80 nm in thickness (see, *e.g.*, Lengeler *et al.* (1999) Biology of Prokaryotes Thieme Verlag: Stuttgart,

p. 913-918, p. 875-899, and p. 88-109 and references therein). The gram-positive cell wall also contains teichoic acid, a polymer of glycerol or ribitol linked through phosphate groups. Teichoic acid is also able to associate with amino acids, and forms covalent bonds with muramic acid. Also present in the cell wall may be lipoteichoic acids and teichuronic acids. If present, cellular surface structures such as flagella or capsules will be anchored in this layer as well.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as HA nucleic acid and protein molecules, which participate in the maintenance of homeostasis in *C. glutamicum*, or which perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the HA molecules participate in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds, in the modification or degradation of aromatic or aliphatic compounds, or have an enzymatic or proteolytic activity. In a preferred embodiment, the activity of the HA molecules of the present invention with regard to *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the HA molecules of the invention are modulated in activity, such that the *C. glutamicum* cellular processes in which the HA molecules participate (*e.g.*, *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity) are also altered in activity, resulting either directly or indirectly in a modulation of the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "HA protein" or "HA polypeptide" includes proteins which participate in a number of cellular processes related to *C. glutamicum* homeostasis or the ability of *C. glutamicum* cells to adapt to unfavorable environmental conditions. For example, an HA protein may be involved in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds in *C. glutamicum*, in the modification or degradation of aromatic or aliphatic compounds in *C. glutamicum*, or have a *C. glutamicum* enzymatic or proteolytic activity. Examples of HA proteins include those encoded by the HA genes set forth in Table 1 and Appendix A. The terms "HA gene" or "HA nucleic acid sequence" include nucleic acid sequences encoding an HA protein, which consist of a coding region and also corresponding untranslated 5' and

3' sequence regions. Examples of HA genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (*i.e.*, fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The term "homeostasis" is art-recognized and includes all of the mechanisms utilized by a cell to maintain a constant intracellular environment despite the prevailing extracellular environmental conditions. A non-limiting example of such processes is the utilization of a cell wall to prevent osmotic lysis due to high intracellular solute concentrations. The term "adaptation" or "adaptation to an environmental condition" is art-recognized and includes mechanisms utilized by the cell to render the cell able to survive under nonpreferred environmental conditions (generally speaking, those environmental conditions in which one or more favored nutrients are absent, or in which an environmental condition such as temperature, pH, osmolarity, oxygen percentage and the like fall outside of the optimal survival range of the cell). Many cells, including *C. glutamicum* cells, possess genes encoding proteins which are expressed under such environmental conditions and which permit continued growth in such suboptimal conditions.

In another embodiment, the HA molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a

microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (*e.g.* phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (*e.g.*, amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of

cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (*e.g.*, regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* HA DNAs and the predicted amino acid sequences of the *C. glutamicum* HA proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins that participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B.

As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, *e.g.*, the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The HA protein or a biologically active portion or fragment thereof of the invention can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections.

A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode HA polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of HA-encoding nucleic acid (*e.g.*, HA DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated HA nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture

medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* HA DNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (*e.g.*, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (*e.g.*, by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (*e.g.*, Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an HA nucleotide sequence can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* HA DNAs of the invention. This DNA comprises sequences encoding HA proteins (*i.e.*, the "coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (*i.e.*, RXA02702, RXN02707, RXS02560, and RXC00110). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA, RXN, RXS, or RXC designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequences in Appendix B designated RXA02702, RXN02707, RXS02560, and RXC00110 are translations of the coding regions of the nucleotide sequence of nucleic acid molecules RXA02702, RXN02707, RXS02560, and RXC00110, respectively, in Appendix A. Each of the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention has also been assigned a SEQ ID NO, as indicated in Table 1.

Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:1, designated, as indicated on Table 1, as "F RXA02702", is an F-designated gene, as are SEQ ID NOs: 9, 11, and 13 (designated on Table 1 as "F RXA02707", "F RXA02708", and "F RXA02709", respectively).

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., *et al.* (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide

sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an HA protein. The nucleotide sequences determined from the cloning of the HA genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning HA homologues in other cell types and organisms, as well as HA homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone HA homologues. Probes based on the HA nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an HA protein, such as by measuring a level of an HA-encoding

nucleic acid in a sample of cells, *e.g.*, detecting HA mRNA levels or determining whether a genomic HA gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. Proteins involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an HA protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of HA protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the HA nucleic acid molecules of the invention are preferably biologically active portions of one of the HA proteins. As used herein, the term "biologically active portion of an HA protein" is intended to include a portion, *e.g.*, a domain/motif, of an HA protein that can participate in the maintenance of homeostasis in *C. glutamicum*, or that can perform a function involved in the adaptation of this microorganism to different environmental conditions, or has an activity as set forth in Table 1. To determine whether an HA protein or a biologically active portion thereof can participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or has a *C. glutamicum* enzymatic or proteolytic activity, an assay

of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an HA protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the HA protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the HA protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same HA protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (*e.g.*, a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 36% identical to the nucleotide sequence designated RXA00009 (SEQ ID NO:85), a nucleotide sequence which is greater than and/or at least 40% identical to the nucleotide sequence designated RXA00277 (SEQ ID NO:91), and a nucleotide sequence which is greater than and/or at least 43% identical to the nucleotide sequence designated RXA00499 (SEQ ID NO:173). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (*e.g.*, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%,

76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

In addition to the *C. glutamicum* HA nucleotide sequences shown in Appendix A, it will be appreciated by those of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of HA proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the HA gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an HA protein, preferably a *C. glutamicum* HA protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the HA gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in HA that are the result of natural variation and that do not alter the functional activity of HA proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* HA DNA of the invention can be isolated based on their homology to the *C. glutamicum* HA nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule

refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* HA protein.

In addition to naturally-occurring variants of the HA sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded HA protein, without altering the functional ability of the HA protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the HA proteins (Appendix B) without altering the activity of said HA protein, whereas an "essential" amino acid residue is required for HA protein activity. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved in the domain having HA activity) may not be essential for activity and thus are likely to be amenable to alteration without altering HA activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HA proteins that contain changes in amino acid residues that are not essential for HA activity. Such HA proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the HA activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participating in the maintenance of homeostasis in *C. glutamicum*, or of performing a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (*e.g.*, one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid

positions or nucleotide positions are then compared. When a position in one sequence (*e.g.*, one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (*e.g.*, a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an HA protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an HA protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an HA coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an HA activity described herein to identify mutants that retain HA activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding HA proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the

coding strand of a double-stranded DNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HA coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an HA protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO. 1 (RXA02702) comprises nucleotides 1 to1458). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding HA. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding HA disclosed herein (*e.g.*, the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of HA mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of HA mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HA mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine,

2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HA protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HA mRNA transcripts to thereby inhibit translation of HA mRNA. A ribozyme having specificity for an HA-encoding nucleic acid can be designed based upon the nucleotide sequence of an HA DNA molecule disclosed herein (*i.e.*, SEQ ID NO. 3 (RXA02705) Appendix A). For example, a derivative of a *Tetrahymena* L-19

IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HA-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071 and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, HA mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, HA gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an HA nucleotide sequence (*e.g.*, an HA promoter and/or enhancers) to form triple helical structures that prevent transcription of an HA gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an HA protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which are

operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, $lacI^q$-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, arny, SPO2, λ-P_R- or λ P_L, which are used preferably in bacteria. Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as ADC1, MFα, AC, P-60, CYC1, GAPDH, TEF, rp28, ADH, promoters from plants such as CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. It is also possible to use artificial promoters. It will be appreciated by those of ordinary skill in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, HA proteins, mutant forms of HA proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of HA proteins in prokaryotic or eukaryotic cells. For example, HA genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. *et al.* (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. *et al.* (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology:*

Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the HA protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant HA protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λgt11, pBdCl, and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89 ; and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter. For transformation

of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming Streptomyces, while plasmids pUB110, pC194, or pBD214 are suited for transformation of Bacillus species. Several plasmids of use in the transfer of genetic information into Corynebacterium include pHM1519, pBL1, pSA77, or pAJ667 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada *et al.* (1992) *Nucleic Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the HA protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), 2 μ, pAG-1, Yep6, Yep13, pEMBLYe23, pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York (IBSN 0 444 904018).

Alternatively, the HA proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the HA proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (*e.g.*, the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for

plant transformation", *Nucl. Acid. Res.* 12: 8711-8721, and include pLGV23, pGHlac+, pBIN19, pAK2004, and pDH51 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HA mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for

instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an HA protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those of ordinary skill in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, linear DNA or RNA (*e.g.*, a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (*e.g.*, a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these

integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an HA protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an HA gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the HA gene. Preferably, this HA gene is a *Corynebacterium glutamicum* HA gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous HA gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous HA gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous HA protein). In the homologous recombination vector, the altered portion of the HA gene is flanked at its 5' and 3' ends by additional nucleic acid of the HA gene to allow for homologous recombination to occur between the exogenous HA gene carried by the vector and an endogenous HA gene in a microorganism. The additional flanking HA nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (*e.g.*, by electroporation) and cells in which the introduced HA gene has homologously recombined with the endogenous HA gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an HA gene on a vector placing it under control of the lac operon permits expression of the HA gene only in the presence of IPTG. Such regulatory systems are well known in the art.

In another embodiment, an endogenous HA gene in a host cell is disrupted (*e.g.*, by homologous recombination or other genetic means known in the art) such that

expression of its protein product does not occur. In another embodiment, an endogenous or introduced HA gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described HA gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present invention.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an HA protein. Accordingly, the invention further provides methods for producing HA proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an HA protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered HA protein) in a suitable medium until HA protein is produced. In another embodiment, the method further comprises isolating HA proteins from the medium or the host cell.

C. Isolated HA Proteins

Another aspect of the invention pertains to isolated HA proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of HA protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of HA protein having less than about 30% (by dry weight) of non-HA protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-HA protein, still more preferably less than about 10% of non-HA protein, and most preferably less than about 5% non-HA protein. When the HA protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes

preparations of HA protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein having less than about 30% (by dry weight) of chemical precursors or non-HA chemicals, more preferably less than about 20% chemical precursors or non-HA chemicals, still more preferably less than about 10% chemical precursors or non-HA chemicals, and most preferably less than about 5% chemical precursors or non-HA chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the HA protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* HA protein in a microorganism such as *C. glutamicum.*

An isolated HA protein or a portion thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (*i.e.*, the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an HA protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% % or more homologous to one of the nucleic acid sequences of Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are

intended to be included. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein. For example, a preferred HA protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or which has one or more of the activities set forth in Table 1.

In other embodiments, the HA protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the HA protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the HA activities described herein. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an HA protein include peptides comprising amino acid sequences derived from the amino acid sequence of an HA protein, *e.g.*, the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an HA protein, which include fewer amino acids than a full length HA protein or the full length protein which is homologous to an HA protein, and exhibit at least one activity of an HA protein. Typically, biologically active portions (peptides, *e.g.*, peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an HA protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an

HA protein include one or more selected domains/motifs or portions thereof having biological activity.

HA proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the HA protein is expressed in the host cell. The HA protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an HA protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native HA protein can be isolated from cells (*e.g.*, endothelial cells), for example using an anti-HA antibody, which can be produced by standard techniques utilizing an HA protein or fragment thereof of this invention.

The invention also provides HA chimeric or fusion proteins. As used herein, an HA "chimeric protein" or "fusion protein" comprises an HA polypeptide operatively linked to a non-HA polypeptide. An "HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an HA protein, whereas a "non-HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the HA protein, *e.g.*, a protein which is different from the HA protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the HA polypeptide and the non-HA polypeptide are fused in-frame to each other. The non-HA polypeptide can be fused to the N-terminus or C-terminus of the HA polypeptide. For example, in one embodiment the fusion protein is a GST-HA fusion protein in which the HA sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HA proteins. In another embodiment, the fusion protein is an HA protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of an HA protein can be increased through use of a heterologous signal sequence.

Preferably, an HA chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor

primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An HA-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HA protein.

Homologues of the HA protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the HA protein. As used herein, the term "homologue" refers to a variant form of the HA protein which acts as an agonist or antagonist of the activity of the HA protein. An agonist of the HA protein can retain substantially the same, or a subset, of the biological activities of the HA protein. An antagonist of the HA protein can inhibit one or more of the activities of the naturally occurring form of the HA protein, by, for example, competitively binding to a downstream or upstream member of a biochemical cascade which includes the HA protein, by binding to a target molecule with which the HA protein interacts, such that no functional interaction is possible, or by binding directly to the HA protein and inhibiting its normal activity.

In an alternative embodiment, homologues of the HA protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the HA protein for HA protein agonist or antagonist activity. In one embodiment, a variegated library of HA variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HA variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HA sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of HA sequences therein. There are a variety of methods which can be used to produce libraries of potential HA homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HA sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the HA protein coding can be used to generate a variegated population of HA fragments for screening and subsequent

selection of homologues of an HA protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an HA coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HA protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HA homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HA homologues (Arkin and Yourvan (1992) *PNAS* *89*:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated HA library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of HA protein regions required for function; modulation of an HA protein activity; modulation of the metabolism of one or more inorganic compounds; modulation of the modification or degradation of one or more aromatic or aliphatic compounds; modulation of cell wall synthesis or rearrangements; modulation of enzyme activity or proteolysis; and modulation of cellular production of a desired compound, such as a fine chemical.

The HA nucleic acid molecules of the invention have a variety of uses. First,

they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at least 5,000 deaths since 1990.

In one embodiment, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in C. *glutamicum* are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic

acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The HA nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The processes involved in adaptation and the maintenance of homeostasis in which the molecules of the invention participate are utilized by a wide variety of species; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the HA nucleic acid molecules of the invention may result in the production of HA proteins having functional differences from the wild-type HA proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more HA proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the HA protein is assessed.

The modulation of activity or number of HA proteins involved in cell wall biosynthesis or rearrangements may impact the production, yield, and/or efficiency of production of one or more fine chemicals from *C. glutamicum* cells. For example, by altering the activity of these proteins, it may be possible to modulate the structure or thickness of the cell wall. The cell wall serves in large measure as a protective device against osmotic lysis and external sources of injury; by modifying the cell wall it may be possible to increase the ability of *C. glutamicum* to withstand the mechanical and shear force stresses encountered by this microorganism during large-scale fermentor culture. Further, each *C. glutamicum* cell is surrounded by a thick cell wall, and thus, a significant portion of the biomass present in large scale culture consists of cell wall. By increasing the rate at which the cell wall is synthesized or by activating cell wall

synthesis (through genetic engineering of the HA cell wall proteins of the invention) it may be possible to improve the growth rate of the microorganism. Similarly, by decreasing the activity or number of proteins involved in the degradation of cell wall or by decreasing the repression of cell wall biosynthesis, an overall increase in cell wall production may be achieved. An increase in the number of viable *C. glutamicum* cells (as may be accomplished by any of the foregoing described protein alterations) should result in increased numbers of cells producing the desired fine chemical in large-scale fermentor culture, which should permit increased yields or efficiency of production of these compounds from the culture.

The modulation of activity or number of *C. glutamicum* HA proteins that participate in the modification or degradation of aromatic or aliphatic compounds may also have direct or indirect impacts on the production of one or more fine chemicals from these cells. Certain aromatic or aliphatic modification or degradation products are desirable fine chemicals (*e.g.*, organic acids or modified aromatic and aliphatic compounds); thus, by modifying the enzymes which perform these modifications (*e.g.*, hydroxylation, methylation, or isomerization) or degradation reactions, it may be possible to increase the yields of these desired compounds. Similarly, by decreasing the activity or number of proteins involved in pathways which further degrade the modified or breakdown products of the aforementioned reactions it may be possible to improve the yields of these fine chemicals from *C. glutamicum* cells in culture.

These aromatic and aliphatic modification and degradative enzymes are themselves fine chemicals. In purified form, these enzymes may be used to degrade aromatic and aliphatic compounds (*e.g.*, toxic chemicals such as petroleum products), either for the bioremediation of polluted sites, for the engineered decomposition of wastes, or for the large-scale and economically feasible production of desired modified aromatic or aliphatic compounds or their breakdown products, some of which may be conveniently used as carbon or energy sources for other fine chemical-producing compounds in culture (see, *e.g.*, Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin and references therein; and Roberts, S.M., ed. (1992-1996) Preparative Biotransformations, Wiley: Chichester, and references therein). By genetically altering these proteins such that their regulation by other cellular mechanisms is lessened or abolished, it may be possible to increase the overall number or activity of these proteins, thereby improving not only the yield of these fine chemicals but also the activity of these harvested proteins.

The modification of these aromatic and aliphatic modifying and degradation enzymes may also have an indirect effect on the production of one or more fine chemical. Many aromatic and aliphatic compounds (such as those that may be

encountered as impurities in culture media or as waste products from cellular metabolism) are toxic to cells; by modifying and/or degrading these compounds such that they may be readily removed or destroyed, cellular viability should be increased. Further, these enzymes may modify or degrade these compounds in such a manner that the resulting products may enter the normal carbon metabolism pathways of the cell, thus rendering the cell able to use these compounds as alternate carbon or energy sources. In large-scale culture situations, when there may be limiting amounts of optimal carbon sources, these enzymes provide a method by which cells may continue to grow and divide using aromatic or aliphatic compounds as nutrients. In either case, the resulting increase in the number of *C. glutamicum* cells in the culture producing the desired fine chemical should in turn result in increased yields or efficiency of production of the fine chemical(s).

Modifications in activity or number of HA proteins involved in the metabolism of inorganic compounds may also directly or indirectly affect the production of one or more fine chemicals from *C. glutamicum* or related bacterial cultures. For example, many desirable fine chemicals, such as nucleic acids, amino acids, cofactors and vitamins (*e.g.*, thiamine, biotin, and lipoic acid) cannot be synthesized without inorganic molecules such as phosphorous, nitrate, sulfate, and iron. The inorganic metabolism proteins of the invention permit the cell to obtain these molecules from a variety of inorganic compounds and to divert them into various fine chemical biosynthetic pathways. Therefore, by increasing the activity or number of enzymes involved in the metabolism of these inorganic compounds, it may be possible to increase the supply of these possibly limiting inorganic molecules, thereby directly increasing the production or efficiency of production of various fine chemicals from *C. glutamicum* cells containing such altered proteins. Modification of the activity or number of inorganic metabolism enzymes of the invention may also render *C. glutamicum* able to better utilize limited inorganic compound supplies, or to utilize nonoptimal inorganic compounds to synthesize amino acids, vitamins, cofactors, or nucleic acids, all of which are necessary for continued growth and replication of the cell. By improving the viability of these cells in large-scale culture, the number of *C. glutamicum* cells producing one or more fine chemicals in the culture may also be increased, in turn increasing the yields or efficiency of production of one or more fine chemicals.

C. glutamicum enzymes for general processes are themselves desirable fine chemicals. The specific properties of enzymes (*i.e.*, regio- and stereospecificity, among others) make them useful catalysts for chemical reactions *in vitro*. Either whole *C. glutamicum* cells may be incubated with an appropriate substrate such that the desired product is produced by enzymes in the cell, or the desired enzymes may be

overproduced and purified from *C. glutamicum* cultures (or those of a related bacterium) and subsequently utilized in *in vitro* reactions in an industrial setting (either in solution or immobilized on a suitable immobile phase). In either situation, the enzyme can either be a natural *C. glutamicum* protein, or it may be mutagenized to have an altered activity; typical industrial uses for such enzymes include as catalysts in the chemical industry (*e.g.*, for synthetic organic chemistry) as food additives, as feed components, for fruit processing, for leather preparation, in detergents, in analysis and medicine, and in the textile industry (see, *e.g.*,Yamada, H. (1993) "Microbial reactions for the production of useful organic compounds," *Chimica* 47: 5-10; Roberts, S.M. (1998) Preparative biotransformations: the employment of enzymes and whole-cells in synthetic chemistry," *J. Chem. Soc. Perkin Trans.* 1: 157-169; Zaks, A. and Dodds, D.R. (1997) "Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals," *DDT* 2: 513-531; Roberts, S.M. and Williamson, N.M. (1997) "The use of enzymes for the preparation of biologically active natural products and analogues in optically active form," *Curr. Organ. Chemistry* 1: 1-20; Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin; Roberts, S.M., ed. (1992-96) Preparative Biotransformations, Wiley: Chichester; Cheetham, P.S.J. (1995) "The applications of enzymes in industry" in : Handbook of Enzyme Biotechnology, 3rd ed., Wiseman, A., ed., Elis: Horwood, p. 419-552; and Ullmann's Encyclopedia of Industrial Chemistry (1987), vol. A9, Enzymes, p. 390-457). Thus, by increasing the activity or number of these enzymes, it may be possible to also increase the ability of the cell to convert supplied substrates to desired products, or to overproduce these enzymes for increased yields in large-scale culture. Further, by mutagenizing these proteins it may be possible to remove feedback inhibition or other repressive cellular regulatory controls such that greater numbers of these enzymes may be produced and activated by the cell, thereby leading to greater yields, production, or efficiency of production of these fine chemical proteins from large-scale cultures. Further, manipulation of these enzymes may alter the activity of one or more *C. glutamicum* metabolic pathways, such as those for the biosynthesis or secretion of one or more fine chemicals.

Mutagenesis of the proteolytic enzymes of the invention such that they are altered in activity or number may also directly or indirectly affect the yield, production, and/or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, by increasing the activity or number of these proteins, it may be possible to increase the ability of the bacterium to survive in large-scale culture, due to an increased ability of the cell to rapidly degrade proteins misfolded in response to the high temperatures, nonoptimal pH, and other stresses encountered during fermentor culture. Increased numbers of cells in these cultures may result in increased yields or efficiency

of production of one or more desired fine chemicals, due to the relatively larger number of cells producing these compounds in the culture. Also, *C. glutamicum* cells possess multiple cell-surface proteases which serve to break down external nutrients into molecules which may be more readily incorporated by the cells as carbon/energy sources or nutrients of other kinds. An increase in activity or number of these enzymes may improve this turnover and increase the levels of available nutrients, thereby improving cell growth or production. Thus, modifications of the proteases of the invention may indirectly impact *C. glutamicum* fine chemical production.

A more direct impact on fine chemical production in response to the modification of one or more of the proteases of the invention may occur when these proteases are involved in the production or degradation of a desired fine chemical. By decreasing the activity of a protease which degrades a fine chemical or a protein involved in the synthesis of a fine chemical it may be possible to increase the levels of that fine chemical (due to the decreased degradation or increased synthesis of the compound). Similarly, by increasing the activity of a protease which degrades a compound to result in a fine chemical or a protein involved in the degradation of a fine chemical, a similar result should be achieved: increased levels of the desired fine chemical from *C. glutamicum* cells containing these engineered proteins.

The aforementioned mutagenesis strategies for HA proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one of ordinary skill in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated HA nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any product produced by *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, Appendices, and the sequence listing cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4 \times 7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4 \times 7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4 \times H_2O$, 10 mg/l $ZnSO_4 \times 7\ H_2O$, 3 mg/l $MnCl_2 \times 4\ H_2O$, 30 mg/l H_3BO_3 20 mg/l $CoCl_2 \times 6\ H_2O$, 1 mg/l $NiCl_2 \times 6\ H_2O$, 3 mg/l $Na_2MoO_4 \times 2\ H_2O$, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20 µg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see e.g.*, Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) *Proc. Natl. Acad. Sci. USA*, 75:3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol* 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) *Gene* 53:283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see *e.g.*, Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (*e.g. Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (*e.g.*, mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as *e.g.*, pHM1519 or pBL1) which replicate autonomously (for review see, *e.g.*,

Martin, J.F. *et al.* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see *e.g.*, Yoshihama, M. *et al.* (1985) *J. Bacteriol.* 162:591-597, Martin J.F. *et al.* (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. *et al.* (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described *e.g.* in Schäfer, A *et al.* (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

Genes may be overexpressed in *C. glutamicum* strains using plasmids which comprise pCG1 (U.S. Patent No. 4,617,267) or fragments thereof, and optionally the gene for kanamycin resistance from TN903 (Grindley, N.D. and Joyce, C.M. (1980) *Proc. Natl. Acad. Sci. USA* 77(12): 7176-7180). In addition, genes may be overexpressed in *C. glutamicum* strains using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

Aside from the use of replicative plasmids, gene overexpression can also be achieved by integration into the genome. Genomic integration in *C. glutamicum* or other Corynebacterium or Brevibacterium species may be accomplished by well-known methods, such as homologous recombination with genomic region(s), restriction

endonuclease mediated integration (REMI) (see, *e.g.*, DE Patent 19823834), or through the use of transposons. It is also possible to modulate the activity of a gene of interest by modifying the regulatory regions (*e.g.*, a promoter, a repressor, and/or an enhancer) by sequence modification, insertion, or deletion using site-directed methods (such as homologous recombination) or methods based on random events (such as transposon mutagenesis or REMI). Nucleic acid sequences which function as transcriptional terminators may also be inserted 3' to the coding region of one or more genes of the invention; such terminators are well-known in the art and are described, for example, in Winnacker, E.L. (1987) From Genes to Clones – Introduction to Gene Technology. VCH: Weinheim.

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. *et al.* (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) *Appl. Microbiol. Biotechnol.*, 32:205-210; von der Osten *et al.* (1998) *Biotechnology Letters*, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if

necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of O.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2,5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well

within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. *et al.* (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (*i.e.*, an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. *et al.*, (1987) "Applications of HPLC in Biochemistry" in:

Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm *et al.* (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. *et al.* (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (*e.g.*, sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum* cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One of ordinary skill in the art would be well-versed in the selection of appropriate

chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek *et al.* (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova *et al.* (1996) *Biotekhnologiya* 11: 27-32; and Schmidt *et al.* (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. *et al.* (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Example 11: Analysis of the Gene Sequences of the Invention

The comparison of sequences and determination of percent homology between two sequences are art-known techniques, and can be accomplished using a mathematical algorithm, such as the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci.* USA 87:2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to HA nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to HA protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, one of ordinary skill in the art will know how to optimize the parameters of the program (*e.g.*, XBLAST and NBLAST) for the specific sequence being analyzed.

Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Meyers and Miller ((1988) *Comput. Appl. Biosci.* 4: 11-17). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art, and include ADVANCE and ADAM. described in Torelli and Robotti (1994) *Comput. Appl. Biosci.* 10:3-5; and FASTA, described in Pearson and Lipman (1988) *P.N.A.S.* 85:2444-8.

The percent homology between two amino acid sequences can also be accomplished using the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. The percent homology between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using standard parameters, such as a gap weight of 50 and a length weight of 3.

A comparative analysis of the gene sequences of the invention with those present in Genbank has been performed using techniques known in the art (see, *e.g.*, Bexevanis and Ouellette, eds. (1998) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley and Sons: New York). The gene sequences of the invention were compared to genes present in Genbank in a three-step process. In a first step, a BLASTN analysis (*e.g.*, a local alignment analysis) was performed for each of the sequences of the invention against the nucleotide sequences present in Genbank, and the top 500 hits were retained for further analysis. A subsequent FASTA search (*e.g.*, a combined local and global alignment analysis, in which limited regions of the sequences are aligned) was performed on these 500 hits. Each gene sequence of the invention was subsequently globally aligned to each of the top three FASTA hits, using the GAP program in the GCG software package (using standard parameters). In order to obtain correct results, the length of the sequences extracted from Genbank were adjusted to the length of the query sequences by methods well-known in the art. The results of this analysis are set forth in Table 4. The resulting data is identical to that which would have been obtained had a GAP (global) analysis alone been performed on each of the genes of the invention in comparison with each of the references in Genbank, but required significantly reduced computational time as compared to such a database-wide GAP (global) analysis. Sequences of the invention for which no alignments above the cutoff values were obtained are indicated on Table 4 by the absence of alignment information. It will further be understood by one of ordinary skill in the art that the GAP alignment

homology percentages set forth in Table 4 under the heading "% homology (GAP)" are listed in the European numerical format, wherein a ',' represents a decimal point. For example, a value of "40,345" in this column represents "40.345%".

Example 12: Construction and Operation of DNA Microarrays

The sequences of the invention may additionally be used in the construction and application of DNA microarrays (the design, methodology, and uses of DNA arrays are well known in the art, and are described, for example, in Schena, M. *et al.* (1995) *Science* 270: 467-470; Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367; DeSaizieu, A. *et al.* (1998) *Nature Biotechnology* 16: 45-48; and DeRisi, J.L. *et al.* (1997) *Science* 278: 680-686).

DNA microarrays are solid or flexible supports consisting of nitrocellulose, nylon, glass, silicone, or other materials. Nucleic acid molecules may be attached to the surface in an ordered manner. After appropriate labeling, other nucleic acids or nucleic acid mixtures can be hybridized to the immobilized nucleic acid molecules, and the label may be used to monitor and measure the individual signal intensities of the hybridized molecules at defined regions. This methodology allows the simultaneous quantification of the relative or absolute amount of all or selected nucleic acids in the applied nucleic acid sample or mixture. DNA microarrays, therefore, permit an analysis of the expression of multiple (as many as 6800 or more) nucleic acids in parallel (see, *e.g.*, Schena, M. (1996) *BioEssays* 18(5): 427-431).

The sequences of the invention may be used to design oligonucleotide primers which are able to amplify defined regions of one or more *C. glutamicum* genes by a nucleic acid amplification reaction such as the polymerase chain reaction. The choice and design of the 5' or 3' oligonucleotide primers or of appropriate linkers allows the covalent attachment of the resulting PCR products to the surface of a support medium described above (and also described, for example, Schena, M. *et al.* (1995) *Science* 270: 467-470).

Nucleic acid microarrays may also be constructed by *in situ* oligonucleotide synthesis as described by Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367. By photolithographic methods, precisely defined regions of the matrix are exposed to light. Protective groups which are photolabile are thereby activated and undergo nucleotide addition, whereas regions that are masked from light do not undergo any modification. Subsequent cycles of protection and light activation permit the synthesis of different oligonucleotides at defined positions. Small, defined regions of the genes of the invention may be synthesized on microarrays by solid phase oligonucleotide synthesis.

The nucleic acid molecules of the invention present in a sample or mixture of nucleotides may be hybridized to the microarrays. These nucleic acid molecules can be labeled according to standard methods. In brief, nucleic acid molecules (*e.g.*, mRNA molecules or DNA molecules) are labeled by the incorporation of isotopically or fluorescently labeled nucleotides, *e.g.*, during reverse transcription or DNA synthesis. Hybridization of labeled nucleic acids to microarrays is described (*e.g.*, in Schena, M. *et al.* (1995) *supra*; Wodicka, L. *et al.* (1997), *supra*; and DeSaizieu A. *et al.* (1998), *supra*). The detection and quantification of the hybridized molecule are tailored to the specific incorporated label. Radioactive labels can be detected, for example, as described in Schena, M. *et al.* (1995) *supra*) and fluorescent labels may be detected, for example, by the method of Shalon *et al.* (1996) *Genome Research* 6: 639-645).

The application of the sequences of the invention to DNA microarray technology, as described above, permits comparative analyses of different strains of *C. glutamicum* or other Corynebacteria. For example, studies of inter-strain variations based on individual transcript profiles and the identification of genes that are important for specific and/or desired strain properties such as pathogenicity, productivity and stress tolerance are facilitated by nucleic acid array methodologies. Also, comparisons of the profile of expression of genes of the invention during the course of a fermentation reaction are possible using nucleic acid array technology.

Example 13: Analysis of the Dynamics of Cellular Protein Populations (Proteomics)

The genes, compositions, and methods of the invention may be applied to study the interactions and dynamics of populations of proteins, termed 'proteomics'. Protein populations of interest include, but are not limited to, the total protein population of *C. glutamicum* (*e.g.*, in comparison with the protein populations of other organisms), those proteins which are active under specific environmental or metabolic conditions (*e.g.*, during fermentation, at high or low temperature, or at high or low pH), or those proteins which are active during specific phases of growth and development.

Protein populations can be analyzed by various well-known techniques, such as gel electrophoresis. Cellular proteins may be obtained, for example, by lysis or extraction, and may be separated from one another using a variety of electrophoretic techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins largely on the basis of their molecular weight. Isoelectric focusing polyacrylamide gel electrophoresis (IEF-PAGE) separates proteins by their isoelectric point (which reflects not only the amino acid sequence but also posttranslational modifications of the protein). Another, more preferred method of protein analysis is the

consecutive combination of both IEF-PAGE and SDS-PAGE, known as 2-D-gel electrophoresis (described, for example, in Hermann *et al.* (1998) *Electrophoresis* 19: 3217-3221; Fountoulakis *et al.* (1998) *Electrophoresis* 19: 1193-1202; Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192; Antelmann *et al.* (1997) *Electrophoresis* 18: 1451-1463). Other separation techniques may also be utilized for protein separation, such as capillary gel electrophoresis; such techniques are well known in the art.

Proteins separated by these methodologies can be visualized by standard techniques, such as by staining or labeling. Suitable stains are known in the art, and include Coomassie Brilliant Blue, silver stain, or fluorescent dyes such as Sypro Ruby (Molecular Probes). The inclusion of radioactively labeled amino acids or other protein precursors (*e.g.*, ^{35}S-methionine, ^{35}S-cysteine, ^{14}C-labelled amino acids, ^{15}N-amino acids, $^{15}NO_3$ or $^{15}NH_4^+$ or ^{13}C-labelled amino acids) in the medium of *C. glutamicum* permits the labeling of proteins from these cells prior to their separation. Similarly, fluorescent labels may be employed. These labeled proteins can be extracted, isolated and separated according to the previously described techniques.

Proteins visualized by these techniques can be further analyzed by measuring the amount of dye or label used. The amount of a given protein can be determined quantitatively using, for example, optical methods and can be compared to the amount of other proteins in the same gel or in other gels. Comparisons of proteins on gels can be made, for example, by optical comparison, by spectroscopy, by image scanning and analysis of gels, or through the use of photographic films and screens. Such techniques are well-known in the art.

To determine the identity of any given protein, direct sequencing or other standard techniques may be employed. For example, N- and/or C-terminal amino acid sequencing (such as Edman degradation) may be used, as may mass spectrometry (in particular MALDI or ESI techniques (see, *e.g.*, Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192)). The protein sequences provided herein can be used for the identification of *C. glutamicum* proteins by these techniques.

The information obtained by these methods can be used to compare patterns of protein presence, activity, or modification between different samples from various biological conditions (*e.g.*, different organisms, time points of fermentation, media conditions, or different biotopes, among others). Data obtained from such experiments alone, or in combination with other techniques, can be used for various applications, such as to compare the behavior of various organisms in a given (*e.g.*, metabolic) situation, to increase the productivity of strains which produce fine chemicals or to increase the efficiency of the production of fine chemicals.

Equivalents

Those of ordinary skill in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

Table 1: Genes in the Application

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
1	2	RXA02702	GR00758	1572	115	UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8)
3	4	RXA02705	GR00758	5803	4388	UDP-N-ACETYLMURAMOYLALANINE--D-GLUTAMATE LIGASE (EC 6.3.2.9)
5	6	RXA01254	GR00365	3807	2539	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE (EC 6.3.2.13)
7	8	RXN02707	VV0017	20110	18581	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D- ALANYL LIGASE
9	10	F RXA02707	GR00758	7264	6920	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D- ALANYL LIGASE (EC 6.3.2.15)
11	12	F RXA02708	GR00758	7694	7260	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D- ALANYL LIGASE (EC 6.3.2.15)
13	14	F RXA02709	GR00758	8451	7723	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D- ALANYL LIGASE (EC 6.3.2.15)
15	16	RXA02710	GR00758	10035	8473	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE (EC 6.3.2.13)
17	18	RXN00531	VV0079	19063	19557	FINE TANGLED PILI MAJOR SUBUNIT
19	20	RXA00944	GR00259	1573	602	NADPH DEHYDROGENASE 3 (EC 1.6.99.1)
21	22	RXS02560	VV0101	9922	10788	NADPH-FLAVIN OXIDOREDUCTASE (EC 1.6.99.-)
23	24	RXS03119	VV0098	86877	87008	SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1)
25	26	RXS03120	VV0098	87351	87476	SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1)

Cell wall biosynthesis

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
27	28	RXA01430	GR00417	7458	6271	N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3.5.1.28)
29	30	RXA02641	GR00749	5097	3022	N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3.5.1.28)
31	32	RXA00135	GR00021	1709	2962	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)
33	34	RXA02706	GR00758	6910	5813	PHOSPHO-N-ACETYLMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13)
35	36	RXA02411	GR00703	1845	997	GLUTAMATE RACEMASE (EC 5.1.1.3)
37	38	RXN01022	VV0143	4460	3381	D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4)
39	40	F RXA01022	GR00292	3	806	D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4)
41	42	RXA02703	GR00758	2698	1610	UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL- UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE (EC 2.4.1.-)
43	44	RXA02711	GR00758	12273	10162	PENICILLIN-BINDING PROTEIN 2

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
45	46	RXA02859	GR10005	846	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)
47	48	RXA00569	GR00152	3928	4953	PENICILLIN-BINDING PROTEIN 4
49	50	RXN03092	VV0054	10445	9561	PENICILLIN-BINDING PROTEIN 1A
51	52	F RXA00594	GR00158	3525	4457	PENICILLIN-BINDING PROTEIN 1A
53	54	RXA01828	GR00516	7736	6315	PENICILLIN-BINDING PROTEIN 3
55	56	RXA00612	GR00162	3	1187	PENICILLIN-BINDING PROTEIN 1A
57	58	RXA01510	GR00424	15370	16650	PENICILLIN-BINDING PROTEIN 4 PRECURSOR (PBP-4) (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4) / D-ALANYL-D-ALANINE-ENDOPEPTIDASE (EC 3.4.99.-)
59	60	RXN01608	VV0139	3536	5374	PENICILLIN-BINDING PROTEIN 5 PRECURSOR
61	62	F RXA01608	GR00449	837	2675	(AL008883) penecillin binding protein [Mycobacterium tuberculosis]
63	64	RXA01270	GR00367	21652	20498	perosamine synthetase
65	66	RXN00549	VV0079	31746	33419	PENICILLIN-BINDING PROTEIN 1A
67	68	RXN00550	VV0079	33457	33777	PENICILLIN-BINDING PROTEIN 1A
69	70	RXN03091	VV0054	9515	8970	PENICILLIN-BINDING PROTEIN 1A
71	72	RXN03178	VV0334	921	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)
73	74	F RXA02859	GR10005	846	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)
75	76	RXN01267	VV0009	17895	16582	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)
77	78	RXN00045	VV0119	4409	5317	UDP-N-acetylglucosamine-2-epimerase (EC 5.1.3.14) /N-acetylmannosamine kinase (EC 2.7.1.60)

Cell division

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
79	80	RXN02704	VV0017	16043	14355	CELL DIVISIN PROTEIN FTSW
81	82	F RXA02704	GR00758	4382	2694	CELL DIVISIN PROTEIN FTSW
83	84	RXA02722	GR00759	2729	1404	CELL DIVISION PROTEIN FTSZ
85	86	RXA00009	GR00002	1545	646	CELL DIVISION PROTEIN FTSX
87	88	RXA00010	GR00002	2248	1562	CELL DIVISION ATP-BINDING PROTEIN FTSE
89	90	RXA00143	GR00022	6328	4847	CELL DIVISION INHIBITOR
91	92	RXA00277	GR00043	1588	5	CELL DIVISION CONTROL PROTEIN 15 (EC 2.7.1.-)
93	94	RXA00857	GR00233	2	1291	CELL DIVISION PROTEIN FTSK
95	96	RXA01435	GR00418	2	871	CELL DIVISION CONTROL PROTEIN 15 (EC 2.7.1.-)
97	98	RXA01511	GR00424	16655	17596	CELL CYCLE PROTEIN MESJ
99	100	RXA01513	GR00424	18368	20926	CELL DIVISION PROTEIN FTSH (EC 3.4.24.-)
101	102	RXA02098	GR00630	4161	5906	CELL DIVISION PROTEIN FTSY
103	104	RXA02713	GR00758	14077	13067	Hypothetical Cell Division Protein mraW
105	106	RXN02723	VV0017	11745	11080	FTSQ
107	108	F RXA02723	GR00759	3460	2984	FTSQ
109	110	RXA01426	GR00417	2777	3403	GLUCOSE INHIBITED DIVISION PROTEIN B

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
111	112	RXA01428	GR00417	4495	5631	STAGE 0 SPORULATION PROTEIN J
113	114	RXA01640	GR00456	4661	1344	STAGE III SPORULATION PROTEIN E
115	116	RXA01829	GR00516	9058	7736	STAGE V SPORULATION PROTEIN E
117	118	RXA01427	GR00417	3512	4432	SOJ PROTEIN
119	120	RXN02973	VV0229	657	4	SOJ PROTEIN
121	122	F RXA01603	GR00447	14043	14663	SOJ PROTEIN
123	124	RXN00818	VV0054	28524	27685	INHIBITION OF MORPHOLOGICAL DIFFERENTIATION

Proteolysis

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
125	126	RXN03028	VV0008	41156	43930	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
127	128	F RXA02470	GR00715	2216	3196	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
129	130	F RXA02471	GR00715	3159	4991	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
131	132	RXN03094	VV0057	1794	43	CLPB PROTEIN
133	134	F RXA01668	GR00464	2205	3920	CLPB PROTEIN
135	136	RXN02937	VV0098	85783	85382	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14)
137	138	RXN03077	VV0043	1729	2913	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14)
139	140	F RXA02855	GR10002	1693	2877	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14), hippurate hydrolase
141	142	RXN00982	VV0149	7596	6091	(L42758) proteinase [Streptomyces lividans]
143	144	F RXA00977	GR00275	1647	2660	(L42758) proteinase [Streptomyces lividans]
145	146	F RXA00982	GR00276	5194	4949	(L42758) proteinase [Streptomyces lividans]
147	148	RXN01181	VV0065	1	957	AMINOPEPTIDASE A/I (EC 3.4.11.1)
149	150	F RXA01181	GR00337	1	957	AMINOPEPTIDASE
151	152	RXN01014	VV0209	13328	10728	AMINOPEPTIDASE N (EC 3.4.11.2)
153	154	F RXA01014	GR00289	3	1580	AMINOPEPTIDASE N (EC 3.4.11.2)
155	156	F RXA01018	GR00290	2289	3152	AMINOPEPTIDASE N (EC 3.4.11.2)
157	158	RXN01046	VV0015	47863	49641	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
159	160	RXN01974	VV0218	3793	5577	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
161	162	RXN01120	VV0182	5678	4401	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX
163	164	F RXA01120	GR00310	2349	1072	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX
165	166	RXN00397	VV0025	3803	4603	XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9)
167	168	RXN01868	VV0127	9980	11905	ZINC METALLOPROTEASE (EC 3.4.24.-)
169	170	F RXA01868	GR00534	1640	30	ZINC METALLOPROTEASE (EC 3.4.24.-)
171	172	F RXA01869	GR00534	1954	1652	ZINC METALLOPROTEASE (EC 3.4.24.-)
173	174	RXN00499	VV0086	8158	9438	PROLINE IMINOPEPTIDASE (EC 3.4.11.5)
175	176	F RXA00499	GR00125	3	959	PROLINE IMINOPEPTIDASE
177	178	RXN01277	VV0009	32155	34158	PROLYL ENDOPEPTIDASE (EC 3.4.21.26)
179	180	F RXA01277	GR00368	1738	50	PROLYL ENDOPEPTIDASE (EC 3.4.21.26)
181	182	RXN00675	VV0005	33258	34049	METHIONINE AMINOPEPTIDASE (EC 3.4.11.18)
183	184	F RXA00675	GR00178	2	484	METHIONINE AMINOPEPTIDASE (EC 3.4.11.18)
185	186	RXN00877	VV0099	2221	3885	PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5)
187	188	F RXA00877	GR00242	3	1067	PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5)

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
189	190	RXN01226	VV0064	4172	4711	PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)
191	192	RXN01963	VV0200	689	6	Hypothetical Secretory Serine Protease (EC 3.4.21.-)
193	194	RXN00621	VV0135	5853	5071	PROTEASE II (EC 3.4.21.83)
195	196	F RXA00621	GR00163	4075	4857	PTRB periplasmic protease
197	198	RXN00622	VV0135	5150	3735	PROTEASE II (EC 3.4.21.83)
199	200	F RXA00622	GR00163	4778	6193	PTRB periplasmic protease
201	202	RXN02146	VV0300	14742	15368	PROTEIN P60 PRECURSOR
203	204	RXN03133	VV0127	39393	40076	HYDROGENASE 1 MATURATION PROTEASE (EC 3.4.-.-)
205	206	RXN02820	VV0131	4799	6109	GAMMA-GLUTAMYLTRANSPEPTIDASE (EC 2.3.2.2)
207	208	F RXA02820	GR00801	1	507	GAMMA-GLUTAMYLTRANSPEPTIDASE (EC 2.3.2.2)
209	210	F RXA02000	GR00589	3430	3933	GAMMA-GLUTAMYLTRANSPEPTIDASE (EC 2.3.2.2)
211	212	RXN02944	VV0169	12751	12074	GAMMA-GLUTAMYLTRANSPEPTIDASE PRECURSOR (EC 2.3.2.2)
213	214	RXS00197	VV0115	2733	1522	Membrane Spanning Protease
215	216	RXS01223	VV0064	7528	8139	PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)
217	218	RXS01642	VV0005	49423	48182	Serine protease

Enzymes in general

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
219	220	RXA01728	GR00489	2452	1478	BETA C-S LYASE (EC 3.-.-.-) PUTATIVE AMINOTRANSFERASE
221	222	RXA02214	GR00650	954	1562	Acetyltransferases
223	224	RXA02716	GR00758	16827	17387	Acetyltransferases
225	226	RXN01499	VV0008	7034	3213	ENTEROBACTIN SYNTHETASE COMPONENT F
227	228	FRXA01499	GR00424	7034	3213	Acetyltransferases (the isoleucine patch superfamily)
229	230	RXN00787	VV0321	3736	5637	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)
231	232	F RXA00787	GR00209	598	5	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)
233	234	F RXA00791	GR00210	831	4	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)
235	236	RXA01057	GR00296	7548	6046	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)
237	238	RXA01055	GR00296	4821	4720	D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1.4.99.1)
239	240	RXA01056	GR00296	5952	5053	D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1.4.99.1)
241	242	RXN02021	VV0160	2008	1061	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase (EC 2.3.1.117)
243	244	RXS00949				quinate dehydrogenase (pyrroloquinoline-quinone) (EC 1.1.99.25)
245	246	RXS00004	VV0196	6930	6460	NITRILASE REGULATOR
247	248	RXS00166	VV0232	3650	4309	Methyltransferase
249	250	RXS00288	VV0079	14586	15596	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)
251	252	RXS01114	VV0182	9118	10341	3-KETOACYL-COA THIOLASE (EC 2.3.1.16)
253	254	RXS01205	VV0268	893	363	UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.-)
255	256	RXS01269	VV0009	21430	20990	UNDECAPRENYL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE (EC 2.7.8.6)
257	258	RXS01421	VV0122	16024	15638	ACYLTRANSFERASE (EC 2.3.1.-)
259	260	RXS01491	VV0139	36800	37450	DNA FOR L-PROLINE 3-HYDROXYLASE, COMPLETE CDS
261	262	RXS01572	VV0009	43945	44436	ALCOHOL DEHYDROGENASE (EC 1.1.1.1)

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
263	264	RXS02453	VV0107	7370	8122	ACETOIN(DIACETYL) REDUCTASE (EC 1.1.1.5)
265	266	RXS02474	VV0008	47021	46248	(S,S)-butane-2,3-diol dehydrogenase (EC 1.1.1.76)
267	268	RXS02485	VV0007	2359	3459	UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE REDUCTASE (EC 1.1.1.158)
269	270	RXS02539	VV0057	17332	15815	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)
271	272	RXS02578	VV0098	7668	6565	ACYLTRANSFERASE
273	274	RXS02741	VV0074	5768	6733	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)
275	276	RXS03061	VV0034	108	437	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)
277	278	RXS03150	VV0155	10678	10055	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)
279	280	RXS02554				Oxidoreductase (EC 1.1.1.-)
281	282	RXS03058				METHYLTRANSFERASE (EC 2.1.1.-)
283	284	RXS03218				CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)
285	286	F RXA01918	GR00549	4644	5057	CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)
287	288	RXC00110	VV0054	27517	26969	Protein involved in hydrolysis of epoxides
289	290	RXC01971	VV0105	4545	3715	Metal-Dependent Hydrolase

Genes encoding enzymes for the metabolism of inorganic compounds

Phosphate and Phosphonate metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
291	292	RXA02118	GR00636	2124	1783	PHNA PROTEIN
293	294	RXA00078	GR00012	6375	5962	PHNB PROTEIN
295	296	RXA02105	GR00632	294	4	PHNB PROTEIN
297	298	RXN00663	VV0142	10120	11493	PHOH PROTEIN HOMOLOG
299	300	F RXA00663	GR00173	1222	227	PHOH PROTEIN HOMOLOG
301	302	RXA00888	GR00242	14325	15341	PHOH PROTEIN HOMOLOG
303	304	RXA01437	GR00418	3932	2550	PHOSPHATE ACETYLTRANSFERASE (EC 2.3.1.8)
305	306	RXN00778	VV0103	18126	19250	PHOSPHATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
307	308	F RXA00778	GR00205	9079	8246	PHOSPHATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
309	310	RXA02497	GR00720	10059	10985	EXOPOLYPHOSPHATASE (EC 3.6.1.11)
311	312	RXA01477	GR00422	8469	10016	ALKALINE PHOSPHATASE D PRECURSOR (EC 3.1.3.1)
313	314	RXA01509	GR00424	15169	14696	INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
315	316	RXA00100	GR00014	9512	10111	DEDA PROTEIN, similar to alkaline phosphatase
317	318	RXA00615	GR00162	3355	2774	DEDA PROTEIN
319	320	RXN00250	VV0189	286	1032	DEDA PROTEIN - ALKALINE PHOSPHATASE LIKE PROTEIN
321	322	F RXA02010	GR00602	79	525	DEDA PROTEIN
323	324	RXA02120	GR00636	5021	4260	CARBOXYVINYL-CARBOXYPHOSPHONATE PHOSPHORYLMUTASE (EC 2.7.8.23)
325	326	RXS01000	VV0106	7252	6407	PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
327	328	RXS01002	VV0106	8858	8055	PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNC
329	330	RXS01003	VV0106	8055	7252	PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
331	332	RXS01902	VV0098	84095	83037	alkaline phosphatase

Fe-Metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
333	334	RXA01967	GR00567	1848	706	FERRIC ENTEROCHELIN ESTERASE HOMOLOG
335	336	RXA00070	GR00011	3436	3867	FERRIC UPTAKE REGULATION PROTEIN
337	338	RXA01934	GR00555	7192	7749	FERRIPYOCHELIN BINDING PROTEIN
339	340	RXN01997	VV0084	33308	33793	FERRITIN
341	342	F RXA01997	GR00586	546	935	FERRITIN
343	344	RXA01082	GR00302	1486	827	IRON REPRESSOR
345	346	RXA01236	GR00358	2185	1241	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
347	348	RXA01354	GR00393	2692	1757	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
349	350	RXA01620	GR00451	2585	3532	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
351	352	RXA02052	GR00624	4586	3795	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
353	354	RXA00372	GR00078	1653	2729	PERIPLASMIC-IRON-BINDING PROTEIN SHIB
355	356	RXA00088	GR00013	4389	5402	FERRIC ANGUIBACTIN-BINDING PROTEIN PRECURSOR
357	358	RXS00156	VV0167	1342	2451	FERROCHELATASE (EC 4.99.1.1)
359	360	RXS00624	VV0135	2018	1332	FERROCHELATASE (EC 4.99.1.1)

Modification and degradation of aromatic compounds

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
361	362	RXA00024	GR00003	938	1882	ARYL-ALCOHOL DEHYDROGENASE (NADP+) (EC 1.1.1.91)
363	364	RXA02526	GR00725	4109	5314	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE (EC 5.5.1.2)
365	366	RXN02813	VV0128	13120	14118	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE HOMOLOG (EC 5.5.1.2)
367	368	F RXA02813	GR00794	651	10	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE HOMOLOG (EC 5.5.1.2)
369	370	RXA01113	GR00307	1098	862	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)
371	372	RXA02126	GR00637	1556	1876	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)
373	374	RXA01465	GR00421	4121	2961	MUCONATE CYCLOISOMERASE (EC 5.5.1.1)
375	376	RXA02316	GR00665	9038	8025	MUCONATE CYCLOISOMERASE (EC 5.5.1.1)
377	378	RXA01464	GR00421	2945	2655	MUCONOLACTONE ISOMERASE (EC 5.3.3.4)
379	380	RXA02603	GR00742	7742	8737	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)
381	382	RXN02839	VV0362	3	449	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)
383	384	F RXA02839	GR00832	3	419	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)
385	386	RXA01502	GR00424	8385	9617	BENZENE 1,2-DIOXYGENASE SYSTEM FERREDOXIN--NAD(+) REDUCTASE COMPONENT (EC 1.18.1.3)
387	388	RXA02828	GR00813	15	572	BIPHENYL-2,3-DIOL 1,2-DIOXYGENASE III (EC 1.13.11.39)
389	390	RXA02064	GR00626	5223	4585	CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)
391	392	RXN00639	VV0128	7858	8712	CATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)
393	394	F RXA00639	GR00168	665	6	CATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)
395	396	RXN01653	VV0321	12867	11407	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A
397	398	F RXA00797	GR00212	445	804	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A
399	400	F RXA01653	GR00458	1909	971	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A
401	402	RXN02530	VV0057	5469	6125	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (EC 1.14.13.8)
403	404	F RXA02530	GR00726	20	469	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (EC 1.14.13.8)
405	406	RXA02083	GR00629	1720	311	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 2 (EC 1.14.13.8)

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig	NT Start	NT Stop	Function
407	408	RXA00892	GR00243	2188	1295	PARANITROBENZYL ESTERASE (EC 3.1.1.-)
409	410	RXA02092	GR00629	12153	10516	PARANITROBENZYL ESTERASE (EC 3.1.1.-)
411	412	RXN00658	VV0083	15705	16397	PHENOL 2-MONOOXYGENASE (EC 1.14.13.7)
413	414	F RXA00658	GR00170	321	4	PHENOL 2 MONOOXYGENASE (EC 1.14.13.7)
415	416	RXA01385	GR00406	5320	3440	PHENOL 2 MONOOXYGENASE (EC 1.14.13.7)
417	418	RXN01461	VV0128	12414	13025	PROTOCATECHUATE 3,4-DIOXYGENASE ALPHA CHAIN (EC 1.13.11.3)
419	420	F RXA01461	GR00421	463	5	PROTOCATECHUATE 3,4-DIOXYGENASE ALPHA CHAIN (EC 1.13.11.3)
421	422	RXA01462	GR00421	1167	478	PROTOCATECHUATE 3,4-DIOXYGENASE BETA CHAIN (EC 1.13.11.3)
423	424	RXN00641	VV0128	7440	5950	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)
425	426	F RXA00640	GR00168	1083	1331	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)
427	428	F RXA00641	GR00168	1533	2573	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)
429	430	RXA00642	GR00168	2616	3107	TOLUATE 1,2-DIOXYGENASE BETA SUBUNIT (EC 1.14.12.-)
431	432	RXA00643	GR00168	3122	4657	TOLUATE 1,2-DIOXYGENASE ELECTRON TRANSFER COMPONENT
433	434	RXN01993	VV0182	16	1143	VANILLATE DEMETHYLASE (EC 1.14.-.-)
435	436	F RXA01993	GR00584	1	366	VANILLATE DEMETHYLASE (EC 1.14.-.-)
437	438	F RXA02012	GR00604	2	670	VANILLATE DEMETHYLASE (EC 1.14.-.-)
439	440	RXA01994	GR00584	373	1347	VANILLATE DEMETHYLASE OXIDOREDUCTASE (EC 1.-.-.-)
441	442	RXA02535	GR00726	6599	7753	XYLENE MONOOXYGENASE ELECTRON TRANSFER COMPONENT
443	444	RXA00964	GR00269	1575	451	1-hydroxy-2-naphthoate 1,2-dioxygenase (EC 1.13.11.38)
445	446	RXN01466	VV0019	7050	6091	ARYLESTERASE (EC 3.1.1.2)
447	448	F RXA01466	GR00422	826	5	ARYLESTERASE (EC 3.1.1.2)
449	450	RXN03036	VV0014	671	6	PROTOCATECHUATE 3,4-DIOXYGENASE BETA CHAIN (EC 1.13.11.3)
451	452	F RXA02895	GR10037	671	6	CHLOROCATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)
453	454	RXA02449	GR00710	1458	2360	hydroxyquinol 1,2-dioxygenase (EC 1.13.11.37)
455	456	RXN00178	VV0174	14670	15554	hydroxyquinol 1,2-dioxygenase (EC 1.13.11.37)
457	458	F RXA00178	GR00028	304	1188	HYDROXYQUINOL-1, 2-DIOXYGENASE
459	460	RXA02111	GR00632	4310	5593	QUINOLINATE SYNTHETASE A
461	462	RXA00644		GR0016 8	4657	CIS-1,2-DIHYDROXYCYCLOHEXA-3,5-DIENE-1-CARBOXYLATE DEHYDROGENASE (EC 1.3.1.55)
463	464	RXN00177	VV0174	13589	14656	MALEYLACETATE REDUCTASE (EC 1.3.1.32)
465	466	F RXA00177	GR00028	3	290	MALEYLACETATE REDUCTASE (EC 1.3.1.32) metabolism of 2,4,5-trichlorophenoxyacetic acid
467	468	RXA02448	GR00710	340	1428	MALEYLACETATE REDUCTASE (EC 1.3.1.32)
469	470	RXA00048	GR00008	2185	527	3-(3-HYDROXYPHENYL) PROPIONATE HYDROXYLASE
471	472	RXA01126	GR00313	2	565	POSSIBLE 2-HYDROXYHEPTA-2,4-DIENE-1,7- DIOATE ISOMAERASE
473	474	RXA01117	GR00309	1713	973	SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (EC 2.8.3.5)
475	476	RXA00772	GR00205	2715	1210	SUCCINYL-COA:COENZYME A TRANSFERASE (EC 2.8.3.-)
477	478	RXA01288	GR00372	2018	1644	SUCCINYL-COA:COENZYME A TRANSFERASE (EC 2.8.3.-)

TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvat corboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 07/20/95
AB003132	murC; ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from coryneform bacteria," *Biochem. Biophys. Res. Commun.*, 236(2):383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from Coryneform bacteria," *Appl. Microbiol. Biotechnol.*, 51(2):223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*," *Biosci. Biotechnol. Biochem.*, 60(10):1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB; gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamolytransferase	
AF036932	aroD	3-dehydroquinate dehydratase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp metabolism," *Microbiology,* 144:1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase; ornithine acetyltransferase; N-acetylglutamate kinase; acetylornithine transminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in Corynebacterium glutamicum," *Mol. Cells.,* 8(3):286-294 (1998)
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the Corynebacterium glutamicum panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli," *Appl. Environ. Microbiol.,* 65(4)1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete[1])	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with Corynebacterium glutamicum," *J. Bacteriol.*, 180(12):3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD; srp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzmye); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in Corynebacterium glutamicum; Isolation of genes involved in biochemical characterization of corresponding proteins," *FEMS Microbiol.*, 173(2):303-310 (1999)
AJ132968	cat	Chloramphenicol aceteyl transferase	
AJ224946	mqo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum," *Eur. J. Biochem.*, 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichtinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of Corynebacterium glutamicum: The channel is formed by a low molecular mass polypeptide," *Biochemistry*, 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Vertes, A.A. et al. "Isolation and characterization of IS31831, a transposable element from Corynebacterium glutamicum," *Mol. Microbiol.*, 11(4):739-746 (1994)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the Corynebacterium glutamicum (Brevibacterium lactofermentum AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," *Microbiology*, 142:3347-3354 (1996)
E01358	hdh; hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL; trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al. "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al. "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04484		Prephenate dehydratase	Sotouchi, N. et al. "Production of L-phenylalanine by fermentation," Patent: JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N. et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93
E05112		Dihydro-dipichorinate synthetase	Hatakeyama, K. et al. "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M. et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent: JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent: JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent: JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomeroreductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent: JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08649		Aspartase	Kohama, K. et al "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al. "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent: JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-trypophan," Patent: JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K. et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of Corynebacterium glutamicum," *J. Bacteriol.*, 174:8065-8072 (1992)
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C. et al. "The cloning and nucleotide sequence of Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," *FEMS Microbiol. Lett.*, 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroreductase	Keilhauer, C. et al. "Isoleucine synthesis in Corynebacterium glutamicum: molecular analysis of the ilvB-ilvN-ilvC operon," *J. Bacteriol.*, 175(17):5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al. "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: expression in Escherichia coli and homology to enzymes II from enteric bacteria," *PNAS USA,* 84(24):8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," *FEMS Microbiol. Lett.,* 119(1-2):137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum," *J. Microbiol. Biotechnol.,* 4(4):256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," *Appl. Environ. Microbiol.,* 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	
L35906	dtxr	Diphtheria toxin repressor	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," *J. Bacteriol.,* 177(2):465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," *J. Bacteriol.,* 167:695-702 (1986)
M16175	5S rRNA		Park, Y-H. et al. "Phylogenetic analysis of the coryneform bacteria by 56 rRNA sequences," *J. Bacteriol.,* 169:1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene,* 52:191-200 (1987)
M16664	trpA	Tryptophan synthase, 3'end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene,* 52:191-200 (1987)
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," *Gene,* 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen. Microbiol.,* 138:1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen. Microbiol.*, 138:1167-1175 (1992)
M89931	aecD; brnQ; yhbw	Beta C-S lyase; branched-chain amino acid uptake carrier; hypothetical protein yhbw	Rossol, I. et al. "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," *J. Bacteriol.*, 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the brnQ gene product," *Arch. Microbiol.*, 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D.M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence," *Appl. Environ. Microbiol.*, 59(3):791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J.P. and Dunican, L.K. (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 tpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR; clgIIR	Putative type II 5-cytosoine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli," *J. Bacteriol.*, 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an McrBC-deficient Escherichia coli strain," *Gene*, 203(2):95-101 (1997)
U14965	recA		
U31224	ppx		Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31230	obg; proB; unkdh	?;gamma glutamyl kinase;similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriiskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of Methylobacillus flagellatum and Corynebacterium glutamicum," *Gene,* 175:15-22 (1996)
U35023	thtR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A Corynebacterium glutamicum gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," *Arch. Microbiol.,* 166(2);76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A Corynebacterium glutamicum gene conferring multidrug resistance in the heterologous host Escherichia coli," *J. Bacteriol.,* 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5''-aminoglycoside phosphotransferase	
U89648		Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the Brevibacterium lactofermentum tryptophan operon," *Nucleic Acids Res.,* 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of Corynebacterium glutamicum and possible mechanisms for modulation of its expression," *Mol. Gen. Genet.,* 212(1):112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: Molecular cloning, nucleotide sequence, and expression," *Mol. Gen. Genet.,* 218(2):330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," *Plant. Mol. Biol.,* 21 (3):487-502 (1993)
X17313	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum fda gene: structural comparison of C. glutamicum fructose-1, 6-biphosphate aldolase to class I and class II aldolases," *Mol. Microbiol.,*
X53993	dapA	L-2, 3-dihydrodipicolinate synthetase (EC 4.2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from Corynebacterium glutamicum," *Nucleic Acids Res.,* 18(21):6421 (1990)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS. Microbiol, Lett.*, 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the Corynebacterium glutamicum lysA gene," *Mol. Microbiol.*, 4(11):1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D.M. et al. "Nucleotide sequence of the Corynebacterium glutamicum trpE gene," *Nucleic Acids Res.*, 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the Corynebacterium glutamicum threonine synthase gene," *Mol. Microbiol.*, 4(10):1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS. Microbiol, Lett.*, 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit; Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from Corynebacterium glutamicum," *Mol. Microbiol.*, 5(5):1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspertate beta-semialdehyde dehydrogenase gene asd in Corynebacterium glutamicum," *Mol. Gen. Genet.*, 224(3):317-324 (1990)
X59403	gap;pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase; triosephosphate isomerase	Eikmanns, B.J. "Identification, sequence analysis, and expression of a Corynebacterium glutamicum gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeras," *J. Bacteriol.*, 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E.R. et al. "Molecular analysis of the Corynebacterium glutamicum gdh gene encoding glutamate dehydrogenase," *Mol. Microbiol.*, 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A.H. et al. "Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake," *Mol. Microbiol.*, 5(12):2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	cop1	Ps1 protein	Joliff, G. et al. "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of Corynebacterium glutamicum: The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," *Mol. Microbiol.*, 6(16):2349-2362 (1992)
X66112	glt	Citrate synthase	Eikmanns, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase," *Microbiol.*, 140:1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	
X69103	csp2	Surface layer protein PS2	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum," *Mol. Microbiol.*, 9(1):97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C. et al. "Identification of IS1206, a Corynebacterium glutamicum IS3-related insertion sequence and phylogenetic analysis," *Mol. Microbiol.*, 14(3):571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in Corynebacterium glutamicum: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," *Appl. Environ. Microbiol.*, 60(1):133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B.J. et al. "Cloning sequence analysis, expression, and inactivation of the Corynebacterium glutamicum icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," *J. Bacteriol.*, 177(3):774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	
X75083, X70584	mtrA	5-methyltryptophan resistance	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of Corynebacterium glutamicum encoding resistance to 5-methyltryptophan," *Biochem. Biophys. Res. Commun.*, 201(3):1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of Corynebacterium glutamicum and Brevibacterium lactofermentum," *Appl. Microbiol. Biotechnol.*, 42(4):575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme," *J. Bacteriol.*, 176(12):3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from Corynebacterium glutamicum," *DNA Seq.*, 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase: sequence analysis," *Microbiology*, 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," *Microbiol.*, 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al. "Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum," *J. Bacteriol.*, 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehrmann, A. et al. "Analysis of different DNA fragments of Corynebacterium glutamicum complementing dapE of Escherichia coli," *Microbiology*, 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R. et al. "Phylogeny of the genus Corynebacterium deduced from analyses of small-subunit ribosomal DNA sequences," *Int. J. Syst. Bacteriol.*, 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences," *Int. J. Syst. Bacteriol.*, 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann, A. et al. "Functional analysis of sequences adjacent to dapE of Corynebacterium glutamicumproline reveals the presence of aroP, which encodes the aromatic amino acid transporter," *J. Bacteriol.*, 177(20):5991-5993 (1995)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X86157	argB; argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway," *Microbiology,* 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase; acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase and acetate kinase," *Microbiology,* 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," *J. Bacteriol.,* 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R.M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of Corynebacterium glutamicum," *J. Biol. Chem.*, 271(10):5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine," *J. Bacteriol.*, 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of Corynebacterium glutamicum, encoding two enzymes involved in L-lysine synthesis," *Biotechnol. Lett.*, 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein; Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum," *Mol. Microbiol.*, 22(5):815-826 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X96580	panB; panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H. et al. "D-pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," *Appl. Environ. Microbiol.*, 65(5):1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer Brevibacterium lactofermentum (Corynebacterium glutamicum ATCC 13869)," *Gene*, 198:217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(9):3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the Corynebacterium glutamicum hom-thrB operon," *Mol. Microbiol.*, 2(1):63-72 (1988)
Y08964	murC; ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase; division initiation protein or cell division protein; cell division protein	Honrubia, M.P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from Brevibacterium lactofermentum," *Mol. Gen. Genet.*, 259(1):97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of Corynebacterium glutamicumproline and characterization of a low-affinity uptake system for compatible solutes," *Arch. Microbiol.*, 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from Corynebacterium glutamicum: characterization, expression and inactivation of the pyc gene," *Microbiology*, 144:915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the leuB gene from Corynebacterium glutamicum," *Appl. Microbiol. Biotechnol.*, 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," *Microbiol.*, 145:539-548 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I," *FEMS Microbiol. Lett.*, 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	
Y18059		Attachment site Corynephage 304L	Moreau, S. et al. "Analysis of the integration functions of φ304L: An integrase module among corynephages," *Virology*, 255(1):150-159 (1999)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	Oguiza, J.A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum: Regulation of argS-lysA cluster expression by arginine," *J. Bacteriol.*,175(22):7356-7362 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," *J. Bacteriol.*, 175(9):2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," *Appl. Environ. Microbiol.*, 60(7)2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	
Z49822	sigA	SigA sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J.A. et al "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dmdR gene," *Gene*, 177:103-107 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869," *Gene*, 170(1):91-94 (1996)

[1] A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
Brevibacterium	ammoniagenes	21054							
Brevibacterium	ammoniagenes	19350							
Brevibacterium	ammoniagenes	19351							
Brevibacterium	ammoniagenes	19352							
Brevibacterium	ammoniagenes	19353							
Brevibacterium	ammoniagenes	19354							
Brevibacterium	ammoniagenes	19355							
Brevibacterium	ammoniagenes	19356							
Brevibacterium	ammoniagenes	21055							
Brevibacterium	ammoniagenes	21077							
Brevibacterium	ammoniagenes	21553							
Brevibacterium	ammoniagenes	21580							
Brevibacterium	ammoniagenes	39101							
Brevibacterium	butanicum	21196							
Brevibacterium	divaricatum	21792	P928						
Brevibacterium	flavum	21474							
Brevibacterium	flavum	21129							
Brevibacterium	flavum	21518							
Brevibacterium	flavum			B11474					
Brevibacterium	flavum			B11472					
Brevibacterium	flavum	21127							
Brevibacterium	flavum	21128							
Brevibacterium	flavum	21427							
Brevibacterium	flavum	21475							
Brevibacterium	flavum	21517							
Brevibacterium	flavum	21528							
Brevibacterium	flavum	21529							

Brevibacterium	flavum			B11477					
Brevibacterium	flavum			B11478					
Brevibacterium	flavum	21127							
Brevibacterium	flavum			B11474					
Brevibacterium	healii	15527							
Brevibacterium	ketoglutamicum	21004							
Brevibacterium	ketoglutamicum	21089							
Brevibacterium	ketosoreductum	21914							
Brevibacterium	lactofermentum				70				
Brevibacterium	lactofermentum				74				
Brevibacterium	lactofermentum				77				
Brevibacterium	lactofermentum	21798							
Brevibacterium	lactofermentum	21799							
Brevibacterium	lactofermentum	21800							
Brevibacterium	lactofermentum	21801							
Brevibacterium	lactofermentum			B11470					
Brevibacterium	lactofermentum			B11471					
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	21420							
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	31269							
Brevibacterium	linens	9174							
Brevibacterium	linens	19391							
Brevibacterium	linens	8377							
Brevibacterium	paraffinolyticum					11160			
Brevibacterium	spec.						717.73		
Brevibacterium	spec.						717.73		
Brevibacterium	spec.	14604							
Brevibacterium	spec.	21860							
Brevibacterium	spec.	21864							
Brevibacterium	spec.	21865							

Brevibacterium	spec.	21866							
Brevibacterium	spec.	19240							
Corynebacterium	acetoacidophilum	21476							
Corynebacterium	acetoacidophilum	13870							
Corynebacterium	acetoglutamicum			B11473					
Corynebacterium	acetoglutamicum			B11475					
Corynebacterium	acetoglutamicum	15806							
Corynebacterium	acetoglutamicum	21491							
Corynebacterium	acetoglutamicum	31270							
Corynebacterium	acetophilum			B3671					
Corynebacterium	ammoniagenes	6872						2399	
Corynebacterium	ammoniagenes	15511							
Corynebacterium	fujiokense	21496							
Corynebacterium	glutamicum	14067							
Corynebacterium	glutamicum	39137							
Corynebacterium	glutamicum	21254							
Corynebacterium	glutamicum	21255							
Corynebacterium	glutamicum	31830							
Corynebacterium	glutamicum	13032							
Corynebacterium	glutamicum	14305							
Corynebacterium	glutamicum	15455							
Corynebacterium	glutamicum	13058							
Corynebacterium	glutamicum	13059							
Corynebacterium	glutamicum	13060							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum	21513							
Corynebacterium	glutamicum	21526							
Corynebacterium	glutamicum	21543							
Corynebacterium	glutamicum	13287							
Corynebacterium	glutamicum	21851							
Corynebacterium	glutamicum	21253							

Corynebacterium	glutamicum	21514							
Corynebacterium	glutamicum	21516							
Corynebacterium	glutamicum	21299							
Corynebacterium	glutamicum	21300							
Corynebacterium	glutamicum	39684							
Corynebacterium	glutamicum	21488							
Corynebacterium	glutamicum	21649							
Corynebacterium	glutamicum	21650							
Corynebacterium	glutamicum	19223							
Corynebacterium	glutamicum	13869							
Corynebacterium	glutamicum	21157							
Corynebacterium	glutamicum	21158							
Corynebacterium	glutamicum	21159							
Corynebacterium	glutamicum	21355							
Corynebacterium	glutamicum	31808							
Corynebacterium	glutamicum	21674							
Corynebacterium	glutamicum	21562							
Corynebacterium	glutamicum	21563							
Corynebacterium	glutamicum	21564							
Corynebacterium	glutamicum	21565							
Corynebacterium	glutamicum	21566							
Corynebacterium	glutamicum	21567							
Corynebacterium	glutamicum	21568							
Corynebacterium	glutamicum	21569							
Corynebacterium	glutamicum	21570							
Corynebacterium	glutamicum	21571							
Corynebacterium	glutamicum	21572							
Corynebacterium	glutamicum	21573							
Corynebacterium	glutamicum	21579							
Corynebacterium	glutamicum	19049							
Corynebacterium	glutamicum	19050							

Corynebacterium	glutamicum	19051							
Corynebacterium	glutamicum	19052							
Corynebacterium	glutamicum	19053							
Corynebacterium	glutamicum	19054							
Corynebacterium	glutamicum	19055							
Corynebacterium	glutamicum	19056							
Corynebacterium	glutamicum	19057							
Corynebacterium	glutamicum	19058							
Corynebacterium	glutamicum	19059							
Corynebacterium	glutamicum	19060							
Corynebacterium	glutamicum	19185							
Corynebacterium	glutamicum	13286							
Corynebacterium	glutamicum	21515							
Corynebacterium	glutamicum	21527							
Corynebacterium	glutamicum	21544							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum			B8183					
Corynebacterium	glutamicum			B8182					
Corynebacterium	glutamicum			B12416					
Corynebacterium	glutamicum			B12417					
Corynebacterium	glutamicum			B12418					
Corynebacterium	glutamicum			B11476					
Corynebacterium	glutamicum	21608							
Corynebacterium	lilium		P973						
Corynebacterium	nitrilophilus	21419				11594			
Corynebacterium	spec.		P4445						
Corynebacterium	spec.		P4446						
Corynebacterium	spec.	31088							
Corynebacterium	spec.	31089							
Corynebacterium	spec.	31090							
Corynebacterium	spec.	31090							

Corynebacterium	spec.	31090							
Corynebacterium	spec.	15954							20145
Corynebacterium	spec.	21857							
Corynebacterium	spec.	21862							
Corynebacterium	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H. et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4^{th} edn), World federation for culture collections world data center on microorganisms, Saimata, Japen.

TABLE 4: ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa00009	1023	GB_IN1:CELZK563	29655	U40061	Caenorhabditis elegans cosmid ZK563.	Caenorhabditis elegans	33,694	9-Nov-95
		GB_IN1:CELZK563	29655	U40061	Caenorhabditis elegans cosmid ZK563.	Caenorhabditis elegans	36,040	9-Nov-95
rxa00010	810	GB_BA1:MTCY164	39150	Z95150	Mycobacterium tuberculosis H37Rv complete genome; segment 135/162.	Mycobacterium tuberculosis	38,442	19-Jun-98
		GB_BA1:MTFTSX	4068	X70031	M.tuberculosis ftsX and ftsE (partial) genes.	Mycobacterium tuberculosis	63,158	06-MAR-1997
		GB_BA1:SHGCPIR	107379	X86780	S.hygroscopicus gene cluster for polyketide immunosuppressant rapamycin.	Streptomyces hygroscopicus	38,875	16-Aug-96
rxa00024	1068	GB_HTG1:CEY113G7_3	110000	AL031113	Caenorhabditis elegans chromosome V clone Y113G7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	36,217	12-Jan-99
		GB_HTG1:CEY113G7_3	110000	AL031113	Caenorhabditis elegans chromosome V clone Y113G7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	36,217	12-Jan-99
		GB_PL2:ATF1C12	111945	AL022224	Arabidopsis thaliana DNA chromosome 4, BAC clone F1C12 (ESSA project).	Arabidopsis thaliana	35,824	20-Sep-99
rxa00048	1782	GB_HTG3:AC008905	129915	AC008905	Homo sapiens chromosome 5 clone CITB-H1_2259I14, *** SEQUENCING IN PROGRESS ***, 40 unordered pieces.	Homo sapiens	38,826	3-Aug-99
		GB_HTG3:AC008905	129915	AC008905	Homo sapiens chromosome 5 clone CITB-H1_2259I14, *** SEQUENCING IN PROGRESS ***, 40 unordered pieces.	Homo sapiens	38,826	3-Aug-99
		GB_HTG3:AC008905	129915	AC008905	Homo sapiens chromosome 5 clone CITB-H1_2259I14, *** SEQUENCING IN PROGRESS ***, 40 unordered pieces.	Homo sapiens	37,379	3-Aug-99
rxa00070	555	GB_BA2:BPEFUR	1003	L31851	Bordetella pertussis DNA repair protein (recN) gene, partial cds; iron regulatory protein (fur) gene, complete cds.	Bordetella pertussis	45,756	17-Apr-95
		GB_BA2:BPU11699	537	U11699	Bordetella pertussis ferric uptake regulator (fur) gene, complete cds.	Bordetella pertussis	47,119	14-Jan-95
		GB_BA1:BTFURRECN	1106	Z48227	B.pertussis fur gene for ferric uptake regulator and partial recN gene.	Bordetella pertussis	45,756	10-Feb-95
rxa00078	537	GB_PR3:HUMCOL2A1Z	31001	L10347	Human pro-alpha1 type II collagen (COL2A1) gene exons 1-54, complete cds.	Homo sapiens	39,010	3-Aug-95
		GB_HTG2:AC006721	135550	AC006721	Caenorhabditis elegans clone Y18H1, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.	Caenorhabditis elegans	40,661	23-Feb-99
		GB_HTG2:AC006721	135550	AC006721	Caenorhabditis elegans clone Y18H1, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.	Caenorhabditis elegans	40,661	23-Feb-99
rxa00088	899	GB_RO:MMCGT6	3009	U48896	Mus musculus UDP-galactose:ceramide galactosyltransferase (Cgt) gene, exon 6 and complete cds.	Mus musculus	35,455	1-Nov-96
		GB_RO:MMCGT6	3009	U48896	Mus musculus UDP-galactose:ceramide galactosyltransferase (Cgt) gene, exon 6 and complete cds.	Mus musculus	34,439	1-Nov-96
rxa00100	723	GB_PL1:CAC41C10	38874	AL033501	C.albicans cosmid Ca41C10.	Candida albicans	36,222	10-Nov-98
		GB_PR4:AC007115	180821	AC007115	Homo sapiens chromosome 12 clone 917O5, complete sequence.	Homo sapiens	33,050	17-Aug-99
		GB_PR4:AC007115	180821	AC007115	Homo sapiens chromosome 12 clone 917O5, complete sequence.	Homo sapiens	34,993	17-Aug-99
rxa00135	1377	GB_BA1:MTCY373	35516	Z73419	Mycobacterium tuberculosis H37Rv complete genome; segment 57/162.	Mycobacterium tuberculosis	60,639	17-Jun-98
		GB_BA1:MLU15186	36241	U15186	Mycobacterium leprae cosmid L471.	Mycobacterium leprae	38,377	09-MAR-1995
		GB_BA1:MTMURAGEN	1257	X96711	M.tuberculosis murA gene.	Mycobacterium tuberculosis	61,575	22-MAR-1996
rxa00143	1605	GB_PAT:I92051	1107	I92051	Sequence 18 from patent US 5726299.	Unknown.	37,773	01-DEC-1998

TABLE 4: ALIGNMENT RESULTS

		GB_PAT:I78761	1107	I78761	Sequence 17 from patent US 5693781.	Unknown.	37,773	3-Apr-98
		GB_BA1:MTCY28	40163	Z95890	Mycobacterium tuberculosis H37Rv complete genome; segment 79/162.	Mycobacterium tuberculosis	36,984	18-Jun-98
rxa00177	1191	GB_GSS14:AQ543786	345	AQ543786	RPCI-11-365L6.TV RPCI-11 Homo sapiens genomic clone RPCI-11-365L6, genomic survey sequence.	Homo sapiens	38,551	19-MAY-1999
		GB_PL2:AF017646	3394	AF017646	Schizosaccharomyces pombe TFIIH subunit p47 (tfh47) gene, complete cds.	Schizosaccharomyces pombe	38,122	17-MAR-1999
		GB_PL1:SPCC1682	37404	AL031525	S.pombe chromosome III cosmid c1682.	Schizosaccharomyces pombe	33,983	14-DEC-1998
rxa00178	1008	GB_BA1:AB016258	2260	AB016258	Arthrobacter sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	Arthrobacter sp.	65,182	8-Sep-99
		GB_BA1:CGPUTP	3791	Y09163	C.glutamicum putP gene.	Corynebacterium glutamicum	38,806	8-Sep-97
		GB_STS:G05495	271	G05495	human STS WI-5918.	Homo sapiens	39,925	8-Jun-95
rxa00277	1684	GB_BA1:MTCY22G10	35420	Z84724	Mycobacterium tuberculosis H37Rv complete genome; segment 21/162.	Mycobacterium tuberculosis	39,976	17-Jun-98
		GB_IN1:CELT03F1	38643	U88169	Caenorhabditis elegans cosmid T03F1.	Caenorhabditis elegans	35,127	7-Feb-97
		GB_IN2:CELK02A2	38261	U23171	Caenorhabditis elegans cosmid K02A2.	Caenorhabditis elegans	36,166	21-MAY-1999
rxa00372	1200	GB_IN2:AC005452	79333	AC005452	Drosophila melanogaster, chromosome 2R, region 43B2-43C2, P1 clone DS07185, complete sequence.	Drosophila melanogaster	37,006	26-Nov-98
		GB_IN2:AC005452	79333	AC005452	Drosophila melanogaster, chromosome 2R, region 43B2-43C2, P1 clone DS07185, complete sequence.	Drosophila melanogaster	34,907	26-Nov-98
		GB_IN1:CELW03F8	34766	AF039041	Caenorhabditis elegans cosmid W03F8.	Caenorhabditis elegans	40,712	1-Jan-98
rxa00389	1683	GB_IN1:AB010703	772	AB010703	Theileria sp. gene for major piroplasm surface protein, partial cds, isolate Kamphaeng Saen.	Theileria sp.	40,285	18-Apr-98
		GB_BA1:LLU08911	619	U08911	Lactobacillus leichmannii putative D-alanine:D-alanine ligase (ddl) gene, partial cds.	Lactobacillus leichmannii	40,194	16-Feb-96
		GB_IN1:TPMS1	822	Z48740	T.parva Tpms1 gene for merozoite surface glycoprotein.	Theileria parva	38,902	15-MAY-1995
rxa00467	792	GB_PR4:DJ293M10	202267	AF111167	Homo sapiens jun dimerization protein gene, partial cds; cfos gene, complete cds; and unknown gene.	Homo sapiens	37,995	7-Apr-99
		GB_PR4:DJ293M10	202267	AF111167	Homo sapiens jun dimerization protein gene, partial cds; cfos gene, complete cds; and unknown gene.	Homo sapiens	36,639	7-Apr-99
		GB_IN1:CEW01C9	21493	Z49969	Caenorhabditis elegans cosmid W01C9, complete sequence.	Caenorhabditis elegans	37,980	23-Nov-98
rxa00499	1404	GB_PR4:AC007206	42732	AC007206	Homo sapiens chromosome 19, cosmid R27370, complete sequence.	Homo sapiens	34,982	4-Apr-99
		GB_EST26:AI344735	462	AI344735	qp05a10.x1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1917114 3' similar to gb:M15800 T-LYMPHOCYTE MATURATION-ASSOCIATED PROTEIN (HUMAN);, mRNA sequence.	Homo sapiens	42,675	2-Feb-99
		GB_PR4:AC006479	161837	AC006479	Homo sapiens clone DJ1051J04, complete sequence.	Homo sapiens	38,462	11-Nov-99
rxa00508	1206	GB_HTG2:AC007111	84245	AC007111	Homo sapiens chromosome 16 clone 1-8F, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.	Homo sapiens	37,931	18-MAR-1999
		GB_HTG2:AC007111	84245	AC007111	Homo sapiens chromosome 16 clone 1-8F, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.	Homo sapiens	37,931	18-MAR-1999
		GB_VI:AF141890	1791	AF141890	Columbid herpesvirus 1 DNA-dependent DNA polymerase gene, partial cds.	columbid herpesvirus 1	39,401	7-Jul-99
rxa00569	1149	GB_PAT:I15213	3728	I15213	Sequence 1 from patent US 5460951.	Unknown.	41,244	2-Apr-96
		GB_PAT:E07353	3728	E07353	cDNA encoding bone-related carboxypeptidase-like protein,OSF-5.	Mus sp.	41,244	29-Sep-97

TABLE 4: ALIGNMENT RESULTS

		GB_HTG1:CEY70G10	152184	AL020987	Caenorhabditis elegans chromosome III clone Y70G10, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	34,148	12-DEC-1997
rxa00612	1077	GB_HTG2:AC005020	177756	AC005020	Homo sapiens clone GS259H13, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Homo sapiens	34,551	12-Jun-98
		GB_HTG2:AC005020	177756	AC005020	Homo sapiens clone GS259H13, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Homo sapiens	34,551	12-Jun-98
		GB_HTG2:AC005020	177756	AC005020	Homo sapiens clone GS259H13, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Homo sapiens	37,628	12-Jun-98
rxa00615	705	GB_GSS15:AQ622921	517	AQ622921	HS_5351_A1_A08_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=927 Col=15 Row=A, genomic survey sequence.	Homo sapiens	38,254	16-Jun-99
		GB_GSS3:B36703	432	B36703	HS-1041-B1-B12-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 823 Col=23 Row=D, genomic survey sequence.	Homo sapiens	44,981	17-OCT-1997
		GB_EST25:AI245926	572	AI245926	qk33c08.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1870766 3' similar to SW:COPG_BOVIN P53620 COATOMER GAMMA SUBUNIT ;, mRNA sequence.	Homo sapiens	38,902	28-Jan-99
rxa00621	906	GB_EST1:D36491	360	D36491	CELK033GYF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk33g11 5', mRNA sequence.	Caenorhabditis elegans	40,390	8-Aug-94
		GB_IN2:CELC16A3	34968	U41534	Caenorhabditis elegans cosmid C16A3.	Caenorhabditis elegans	35,477	18-MAY-1999
		GB_HTG3:AC009311	160198	AC009311	Homo sapiens clone NH0311L03, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,636	13-Aug-99
rxa00622	1539	GB_BA1:AB004795	3039	AB004795	Pseudomonas sp. gene for dipeptidyl aminopeptidase, complete cds.	Pseudomonas sp.	54,721	5-Feb-99
		GB_BA1:MBOPII	2392	D38405	Moraxella lacunata gene for protease II, complete cds.	Moraxella lacunata	50,167	8-Feb-99
		GB_IN2:AF078916	2960	AF078916	Trypanosoma brucei brucei oligopeptidase B (opb) gene, complete cds.	Trypanosoma brucei brucei	48,076	08-OCT-1999
rxa00639	978	GB_BA2:AF043741	1223	AF043741	Rhodococcus rhodochrous catechol 1,2-dioxygenase (catA) gene, complete cds.	Rhodococcus rhodochrous	66,940	27-Aug-98
		GB_BA1:D83237	1626	D83237	Rhodococcus erythropolis DNA for catechol 1,2-dioxgenase, complete cds.	Rhodococcus erythropolis	65,440	1-Sep-99
		GB_BA1:ROX99622	7224	X99622	Rhodococcus opacus catR, catA, catB, catC genes and five ORFs.	Rhodococcus opacus	63,617	24-Sep-97
rxa00641	1614	GB_BA2:AF134348	5000	AF134348	Pseudomonas putida plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	Pseudomonas putida	59,863	20-MAY-1999
		GB_BA1:PWWXYL	9037	M64747	Pseudomonas putida plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	59,588	26-Apr-93
		GB_BA1:PCCBDABC	3548	X79076	P.cepacia (2CBS) cbdA, cbdB and cbdC genes.	Burkholderia cepacia	55,410	3-Apr-97
rxa00642	615	GB_BA2:AF134348	5000	AF134348	Pseudomonas putida plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	Pseudomonas putida	60,920	20-MAY-1999
		GB_BA1:PWWXYL	9037	M64747	Pseudomonas putida plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	58,756	26-Apr-93
		GB_GSS11:AQ274007	637	AQ274007	nbxb0032I07f CUGI Rice BAC Library Oryza sativa genomic clone nbxb0032I07f, genomic survey sequence.	Oryza sativa	41,390	3-Nov-98

TABLE 4: ALIGNMENT RESULTS

rxa00643	1659	GB_BA2:AF134348	5000	AF134348	Pseudomonas putida plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	Pseudomonas putida	53,871	20-MAY-1999
		GB_BA1:PWWXYL	9037	M64747	Pseudomonas putida plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	52,603	26-Apr-93
		GB_EST22:AI020666	328	AI020666	ua97f07.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone IMAGE:1365445 5' similar to SW:DUS7_RAT Q63340 DUAL SPECIFICITY PROTEIN PHOSPHATASE 7 ;, mRNA sequence.	Mus musculus	43,865	16-Jun-98
rxa00644	951	GB_BA1:PWWXYL	9037	M64747	Pseudomonas putida plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	55,626	26-Apr-93
		GB_BA2:AF134348	5000	AF134348	Pseudomonas putida plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	Pseudomonas putida	50,410	20-MAY-1999
		GB_EST22:AI038396	438	AI038396	ox21g10.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1657026 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;, mRNA sequence.	Homo sapiens	40,138	28-Aug-98
rxa00658	816	GB_EST16:C26090	414	C26090	C26090 Rice callus cDNA Oryza sativa cDNA clone C11617_1A, mRNA sequence.	Oryza sativa	40,636	6-Aug-97
		GB_EST16:C26090	414	C26090	C26090 Rice callus cDNA Oryza sativa cDNA clone C11617_1A, mRNA sequence.	Oryza sativa	38,406	6-Aug-97
rxa00663	1497	GB_BA1:MTV017	67200	AL021897	Mycobacterium tuberculosis H37Rv complete genome; segment 48/162.	Mycobacterium tuberculosis	57,976	24-Jun-99
		GB_BA1:MLCB1222	34714	AL049491	Mycobacterium leprae cosmid B1222.	Mycobacterium leprae	39,669	27-Aug-99
		GB_HTG2:AC007482	155357	AC007482	Homo sapiens clone hRPK.56_A_1, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.	Homo sapiens	36,154	05-MAY-1999
rxa00675	915	GB_BA1:SC3C8	33095	AL023861	Streptomyces coelicolor cosmid 3C8.	Streptomyces coelicolor	36,836	15-Jan-99
		GB_PR3:AC005736	215441	AC005736	Homo sapiens chromosome 16, BAC clone 462G18 (LANL), complete sequence.	Homo sapiens	42,027	01-OCT-1998
		GB_IN2:AC005719	188357	AC005719	Drosophila melanogaster, chromosome 2L, region 38A5-38B4, BAC clone BACR48M05, complete sequence.	Drosophila melanogaster	35,531	27-OCT-1999
rxa00762	999	GB_HTG2:HSJ473J16	203460	AL109942	Homo sapiens chromosome 6 clone RP3-473J16 map q25.3-26, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,295	03-DEC-1999
		GB_HTG2:HSJ473J16	203460	AL109942	Homo sapiens chromosome 6 clone RP3-473J16 map q25.3-26, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,295	03-DEC-1999
		GB_PR2:HSU91327	129252	U91327	Human chromosome 12p15 BAC clone CIT987SK-99D8 complete sequence.	Homo sapiens	35,650	21-Aug-97
rxa00772	1629	GB_BA2:AF010184	1494	AF010184	Pseudomonas aeruginosa coenzyme A transferase PsecoA (psecoA) gene, complete cds.	Pseudomonas aeruginosa	56,472	18-Jul-98
		GB_PAT:I92043	713	I92043	Sequence 10 from patent US 5726299.	Unknown.	92,701	01-DEC-1998
		GB_PAT:I78754	713	I78754	Sequence 10 from patent US 5693781.	Unknown.	92,701	3-Apr-98
rxa00778	1248	GB_BA1:MTPST2GN	1347	Z48056	M.tuberculosis PstS-2 gene.	Mycobacterium tuberculosis	47,791	24-Apr-99
		GB_BA1:D90907	132419	D90907	Synechocystis sp. PCC6803 complete genome, 9/27, 1056467-1188885.	Synechocystis sp.	35,536	7-Feb-99
		GB_BA1:D90907	132419	D90907	Synechocystis sp. PCC6803 complete genome, 9/27, 1056467-1188885.	Synechocystis sp.	38,006	7-Feb-99

TABLE 4: ALIGNMENT RESULTS

rxa00787	2025	GB_PL1:SCX11RA	36849	X91258	S.cerevisiae DNA from chromosome XII right arm including ACE2, CKI1, PDC5, SLS1, PUT1 and tRNA-Asp genes.	Saccharomyces cerevisiae	36,122	13-OCT-1995
		GB_PL2:YSCL9606	29154	U53881	Saccharomyces cerevisiae chromosome XII cosmid 9606.	Saccharomyces cerevisiae	36,122	25-OCT-1997
		GB_PL1:SCX11RA	36849	X91258	S.cerevisiae DNA from chromosome XII right arm including ACE2, CKI1, PDC5, SLS1, PUT1 and tRNA-Asp genes.	Saccharomyces cerevisiae	37,198	13-OCT-1995
rxa00792	1320	GB_PR4:AC004841	132072	AC004841	Homo sapiens PAC clone DJ0607J23 from 7q21.2-q31.1, complete sequence.	Homo sapiens	37,452	18-MAR-1999
		GB_HTG2:AC006706	180664	AC006706	Caenorhabditis elegans clone Y110A2, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Caenorhabditis elegans	34,824	23-Feb-99
		GB_HTG2:AC006706	180664	AC006706	Caenorhabditis elegans clone Y110A2, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Caenorhabditis elegans	34,824	23-Feb-99
rxa00857	1313	GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	38,080	17-Jun-98
		GB_BA1:MSGY154	40221	AD000002	Mycobacterium tuberculosis sequence from clone y154.	Mycobacterium tuberculosis	68,345	03-DEC-1996
		GB_BA1:MLCB33	42224	Z94723	Mycobacterium leprae cosmid B33.	Mycobacterium leprae	38,824	24-Jun-97
rxa00877	1788	GB_PAT:I92050	567	I92050	Sequence 17 from patent US 5726299.	Unknown.	62,787	01-DEC-1998
		GB_PAT:I78760	567	I78760	Sequence 16 from patent US 5693781.	Unknown.	62,787	3-Apr-98
		GB_BA2:AE000426	10240	AE000426	Escherichia coli K-12 MG1655 section 316 of 400 of the complete genome.	Escherichia coli	36,456	12-Nov-98
rxa00888	1140	GB_BA1:MTCY27	27548	Z95208	Mycobacterium tuberculosis H37Rv complete genome; segment 104/162.	Mycobacterium tuberculosis	40,165	17-Jun-98
		GB_BA1:U00016	42931	U00016	Mycobacterium leprae cosmid B1937.	Mycobacterium leprae	58,444	01-MAR-1994
		GB_BA1:ECU82598	136742	U82598	Escherichia coli genomic sequence of minutes 9 to 12.	Escherichia coli	37,876	15-Jan-97
rxa00892	1017	GB_BA2:AE000817	13157	AE000817	Methanobacterium thermoautotrophicum from bases 251486 to 264642 (section 23 of 148) of the complete genome.	Methanobacterium thermoautotrophicum	36,710	15-Nov-97
		GB_EST29:AI620549	239	AI620549	tu95b07.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2258773 3' similar to gb:X60708_rna1 DIPEPTIDYL PEPTIDASE IV (HUMAN);, mRNA sequence.	Homo sapiens	38,075	21-Apr-99
		GB_BA2:AE000817	13157	AE000817	Methanobacterium thermoautotrophicum from bases 251486 to 264642 (section 23 of 148) of the complete genome.	Methanobacterium thermoautotrophicum	35,650	15-Nov-97
rxa00897	1128	GB_PR3:HS246D7	28011	AL031843	Human DNA sequence from clone 246D7 on chromosome 22q13.1-13.33. Contains ESTs, a GSS and an STS, complete sequence.	Homo sapiens	38,724	23-Nov-99
		GB_PR3:HSDJ185D5	24387	AL118498	Human DNA sequence from clone 185D5 on chromosome 22, complete sequence.	Homo sapiens	37,021	23-Nov-99
		GB_PR3:HS246D7	28011	AL031843	Human DNA sequence from clone 246D7 on chromosome 22q13.1-13.33. Contains ESTs, a GSS and an STS, complete sequence.	Homo sapiens	36,054	23-Nov-99
rxa00944	1095	GB_BA1:ECU68759	1531	U68759	Enterobacter cloacae pentaerythritol tetranitrate reductase (onr) gene, complete cds.	Enterobacter cloacae	43,041	14-DEC-1996
		GB_PAT:A59288	1531	A59288	Sequence 1 from Patent WO9703201.	unidentified	43,041	06-MAR-1998
		GB_EST23:AI099394	601	AI099394	ue32a09.y1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1482040 5' similar to gb:U21301 Mus musculus c-mer tyrosine kinase receptor mRNA, complete (MOUSE);, mRNA sequence.	Mus musculus	37,225	20-Aug-98
rxa00964	1248	GB_HTG6:AC009794	152794	AC009794	Homo sapiens chromosome 4 clone RP11-343C10 map 4, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	Homo sapiens	34,762	03-DEC-1999
		GB_HTG6:AC009794	152794	AC009794	Homo sapiens chromosome 4 clone RP11-343C10 map 4, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	Homo sapiens	35,708	03-DEC-1999

TABLE 4: ALIGNMENT RESULTS

rxa00982	1629	GB_BA1:BLARGS	2501	Z21501	B.lactofermentum argS and lysA genes for arginyl-tRNA synthetase and diaminopimelate decarboxylase (partial).	Corynebacterium glutamicum	39,003	28-DEC-1993
		GB_BA1:CGXLYSA	2344	X54740	Corynebacterium glutamicum argS-lysA operon gene for the upstream region of the arginyl-tRNA synthetase and diaminopimelate decarboxylase (EC 4.1.1.20).	Corynebacterium glutamicum	41,435	30-Jun-93
		GB_PAT:E14508	3579	E14508	DNA encoding Brevibacterium diaminopimelic acid decarboxylase and arginyl-tRNA synthase.	Corynebacterium glutamicum	40,566	28-Jul-99
rxa01014	2724	GB_BA1:MTV008	63033	AL021246	Mycobacterium tuberculosis H37Rv complete genome; segment 108/162.	Mycobacterium tuberculosis	56,167	17-Jun-98
		GB_BA1:STMAMPEPN	2849	L23172	Streptomyces lividans aminopeptidase N gene, complete cds.	Streptomyces lividans	57,067	18-MAY-1994
		GB_BA1:SC7H2	42655	AL109732	Streptomyces coelicolor cosmid 7H2.	Streptomyces coelicolor A3(2)	37,551	2-Aug-99
rxa01022	1203	GB_PAT:A68384	1080	A68384	Sequence 1 from Patent WO9748809.	Mycobacterium avium	56,913	06-MAY-1999
		GB_BA2:AF077728	1346	AF077728	Mycobacterium smegmatis D-alanine:D-alanine ligase gene, complete cds.	Mycobacterium smegmatis	57,203	1-Jan-99
rxa01055		GB_BA1:MSGB1723CS	38477	L78825	Mycobacterium leprae cosmid B1723 DNA sequence.	Mycobacterium leprae	54,599	15-Jun-96
rxa01056	1023	GB_BA2:AE001715	11086	AE001715	Thermotoga maritima section 27 of 136 of the complete genome.	Thermotoga maritima	39,034	2-Jun-99
		GB_EST38:AW046857	161	AW046857	UI-M-BH1-akl-a-04-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone UI-M-BH1-akl-a-04-0-UI 3', mRNA sequence.	Mus musculus	45,963	18-Sep-99
		GB_EST38:AW049435	244	AW049435	UI-M-BH1-ams-b-01-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone UI-M-BH1-ams-b-01-0-UI 3', mRNA sequence.	Mus musculus	40,984	18-Sep-99
rxa01057	1626	GB_PL1:LPAJ5046	656	AJ225046	Lycopersicon peruvianum mRNA for Hsp20.1 protein.	Lycopersicon peruvianum	37,117	22-Jul-98
		GB_PL2:SPAC806	22870	AL117212	S.pombe chromosome I cosmid c806.	Schizosaccharomyces pombe	38,211	24-Nov-99
		GB_PL2:SPAC806	22870	AL117212	S.pombe chromosome I cosmid c806.	Schizosaccharomyces pombe	36,934	24-Nov-99
rxa01082	783	GB_BA2:AF112535	4363	AF112535	Corynebacterium glutamicum putative glutaredoxin NrdH (nrdH), NrdI (nrdI), and ribonucleotide reductase alpha-chain (nrdE) genes, complete cds.	Corynebacterium glutamicum	99,794	5-Aug-99
		GB_PL2:TAE237897	8020	AJ237897	Triticum aestivum sbe1 gene, exons 1-14.	Triticum aestivum	37,132	1-Nov-99
		GB_PL2:AF076680	10499	AF076680	Aegilops tauschii starch branching enzyme-I (SBE-I) gene, complete cds.	Aegilops tauschii	38,651	14-MAY-1999
rxa01113	260	GB_VI:ASU02468	11424	U02468	African swine fever virus BA71V (A489R, A280R, A505R, A498R, A528R, A506R, and A542R) genes, complete cds.	African swine fever virus	31,923	28-Apr-94
		GB_VI:ASU18466	170101	U18466	African swine fever virus, complete genome.	African swine fever virus	31,923	22-Apr-95
		GB_GSS5:AQ752779	1647	AQ752779	HS_5569_B1_D02_SP6 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1145 Col=3 Row=H, genomic survey sequence.	Homo sapiens	37,154	19-Jul-99
rxa01115	876	GB_BA1:AB014757	6057	AB014757	Pseudomonas sp. 61-3 genes for PhbR, acetoacetyl-CoA reductase, beta-ketothiolase and PHB synthase, complete cds.	Pseudomonas sp. 61-3	40,850	26-DEC-1998
		GB_IN2:DMU60591	5630	U60591	Drosophila melanogaster kuzbanian (kuz) mRNA, complete cds.	Drosophila melanogaster	37,326	10-Sep-96
		GB_RO:MMMMP10	1744	Y13185	Mus musculus mRNA for stromelysin-2.	Mus musculus	35,877	14-Jan-98
rxa01116	735	GB_BA1:SC4C6	30941	AL079355	Streptomyces coelicolor cosmid 4C6.	Streptomyces coelicolor	40,616	21-Jun-99
		GB_BA2:AF109386	6551	AF109386	Streptomyces sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	Streptomyces sp. 2065	64,099	06-DEC-1999
		GB_BA1:MTCY07A7	23967	Z95556	Mycobacterium tuberculosis H37Rv complete genome; segment 109/162.	Mycobacterium tuberculosis	41,716	17-Jun-98

TABLE 4: ALIGNMENT RESULTS

rxa01117	864	GB_BA2:AF109386	6551	AF109386	Streptomyces sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	Streptomyces sp. 2065	62,116	06-DEC-1999
		GB_BA2:AF003947	5475	AF003947	Rhodococcus opacus succinyl CoA:3-oxoadipate CoA transferase subunit homolog (pcaI') gene, partial cds, protocatechuate dioxygenase beta subunit (pcaH), protocatechuate dioxygenase alpha subunit (pcaG), 3-carboxy-cis,cis-muconate cycloisomerase homolog (pcaB), 3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase (pcaL) and PcaR (pcaR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcaF') gene, partial cds.	Rhodococcus opacus	36,712	12-MAR-1998
		GB_BA1:XCLPSIJ	2578	Y11313	X.campestris lpsI, lpsJ, xanA genes and orfX.	Xanthomonas campestris	39,833	20-Jan-98
rxa01120	1401	GB_BA1:MTV008	63033	AL021246	Mycobacterium tuberculosis H37Rv complete genome; segment 108/162.	Mycobacterium tuberculosis	36,715	17-Jun-98
		GB_BA1:CAJ10321	6710	AJ010321	Caulobacter crescentus partial tig gene and clpP, cicA, clpX, lon genes.	Caulobacter crescentus	63,311	01-OCT-1998
		GB_BA2:AF150957	4440	AF150957	Azospirillum brasilense trigger factor (tig), heat-shock protein ClpP (clpP), and heat-shock protein ClpX (clpX) genes, complete cds; and Lon protease (lon) gene, partial cds.	Azospirillum brasilense	60,613	7-Jun-99
rxa01126	583	GB_HTG3:AC009199	66498	AC009199	Drosophila melanogaster chromosome 2 clone BACR10J23 (D1024) RPCI-98 10.J.23 map 37B-37B strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 79 unordered pieces.	Drosophila melanogaster	35,294	20-Sep-99
		GB_HTG3:AC009199	66498	AC009199	Drosophila melanogaster chromosome 2 clone BACR10J23 (D1024) RPCI-98 10.J.23 map 37B-37B strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 79 unordered pieces.	Drosophila melanogaster	35,294	20-Sep-99
		GB_PL1:AB016880	81284	AB016880	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MTG10, complete sequence.	Arabidopsis thaliana	34,477	20-Nov-99
rxa01181	980	GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	61,570	22-Aug-97
		GB_BA1:MTCY190	34150	Z70283	Mycobacterium tuberculosis H37Rv complete genome; segment 98/162.	Mycobacterium tuberculosis	60,434	17-Jun-98
		GB_BA1:SC5F7	40024	AL096872	Streptomyces coelicolor cosmid 5F7.	Streptomyces coelicolor A3(2)	57,011	22-Jul-99
rxa01236	1068	GB_EST3:H01832	381	H01832	yj28c11.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:150068 3', mRNA sequence.	Homo sapiens	41,406	19-Jun-95
		GB_PR4:AC004850	105891	AC004850	Homo sapiens PAC clone DJ0665C04 from 7p14-p13, complete sequence.	Homo sapiens	37,428	26-Feb-99
		GB_GSS11:AQ304150	528	AQ304150	HS_3208_A1_D12_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3208 Col=23 Row=G, genomic survey sequence.	Homo sapiens	37,421	16-DEC-1998
rxa01254	1392	GB_BA1:MTV025	121125	AL022121	Mycobacterium tuberculosis H37Rv complete genome; segment 155/162.	Mycobacterium tuberculosis	58,315	24-Jun-99
		GB_BA1:MSGB577COS	37770	L01263	M. leprae genomic dna sequence, cosmid b577.	Mycobacterium leprae	56,323	14-Jun-96
		GB_BA1:MLCB2407	35615	AL023596	Mycobacterium leprae cosmid B2407.	Mycobacterium leprae	37,645	27-Aug-99
rxa01270	1278	GB_BA1:BSPX91182	345	X91182	Bacterial sp. partial 16S rRNA gene (clone group G10).	unidentified bacterium	41,228	15-Jul-96
		GB_BA1:BSPJN12D	347	Z69277	Bacterial sp. partial 16S rRNA gene (clone group JN12d).	Bacteria	38,905	24-Jun-98
		GB_EST7:W93397	545	W93397	zd95b03.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:357197 3', mRNA sequence.	Homo sapiens	40,516	25-Nov-96
rxa01277	2127	GB_PL2:AF111709	52684	AF111709	Oryza sativa subsp. indica Retrosat 1 retrotransposon and Ty3-Gypsy type Retrosat 2 retrotransposon, complete sequences; and unknown genes.	Oryza sativa subsp. indica	37,410	26-Apr-99
		GB_IN1:CELZC250	34372	AF003383	Caenorhabditis elegans cosmid ZC250.	Caenorhabditis elegans	35,506	14-MAY-1997

TABLE 4: ALIGNMENT RESULTS

		GB_EST1:Z14808	331	Z14808	CEL5E4 Chris Martin sorted cDNA library Caenorhabditis elegans cDNA clone cm5e4 5', mRNA sequence.	Caenorhabditis elegans	36,890	19-Jun-97
rxa01288	498	GB_VI:S62819	3348	S62819	F2L=putative RNA polymerase-associated transcription factor. . .F4R=type I topoisomerase homolog [orf virus OV, NZ2, host=sheep, Genomic, 3 genes, 3348 nt].	orf virus	40,471	25-Aug-93
		GB_PR4:HUMCCLEC1	17079	AF077344	Homo sapiens cartilage-derived C-type lectin (CLECSF1) gene, exons 1 and 2.	Homo sapiens	34,631	15-OCT-1999
		GB_PR4:HUMCCLEC1	17079	AF077344	Homo sapiens cartilage-derived C-type lectin (CLECSF1) gene, exons 1 and 2.	Homo sapiens	39,300	15-OCT-1999
rxa01354	1059	GB_PR1:D87675	301692	D87675	Homo sapiens DNA for amyloid precursor protein, complete cds.	Homo sapiens	37,984	22-Sep-97
		GB_PR1:D87675	301692	D87675	Homo sapiens DNA for amyloid precursor protein, complete cds.	Homo sapiens	35,140	22-Sep-97
		GB_RO:MMNUCLEO	11478	X07699	Mouse nucleolin gene.	Mus musculus	37,146	27-Aug-98
rxa01376	984	GB_BA1:MTCY71	42729	Z92771	Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.	Mycobacterium tuberculosis	39,496	10-Feb-99
		GB_BA1:ACCPSXM	2748	X81320	A.calcoaceticus epsX and epsM genes.	Acinetobacter calcoaceticus	40,353	19-OCT-1994
		GB_BA2:ECU05248	1781	U05248	Escherichia coli polysialic acid gene cluster region 2 (neuD and neuB) genes, complete cds.	Escherichia coli	34,995	1-Feb-95
rxa01385	2004	GB_BA1:FVBPENTA	2519	M98557	Flavobacterium sp. pentachlorophenol 4-monooxygenase gene, complete mRNA.	Flavobacterium sp.	40,855	26-Apr-93
		GB_PAT:I19994	2516	I19994	Sequence 2 from patent US 5512478.	Unknown.	40,855	07-OCT-1996
		GB_BA2:AF059680	2410	AF059680	Sphingomonas sp. UG30 pentachlorophenol 4-monooxygenase (pcpB) gene, complete cds; and pentachlorophenol 4-monooxygenase reductase (pcpD) gene, partial cds.	Sphingomonas sp. UG30	42,993	27-Apr-99
rxa01426	750	GB_GSS3:B35912	313	B35912	HS-1031-A2-D02-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=4 Row=G, genomic survey sequence.	Homo sapiens	38,019	17-OCT-1997
		GB_GSS1:FR0027767	497	AL020589	F.rubripes GSS sequence, clone 197B17aA3, genomic survey sequence.	Fugu rubripes	35,814	10-DEC-1997
		GB_GSS5:AQ774340	449	AQ774340	HS_3137_A2_E11_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3137 Col=22 Row=I, genomic survey sequence.	Homo sapiens	40,535	29-Jul-99
rxa01427	1044	GB_BA2:AF036766	3487	AF036766	Lactobacillus reuteri plasmid pTE15 replication-associated protein A (repA) and replication-associated protein B (repB) genes, complete cds.	Lactobacillus reuteri	39,101	19-Feb-98
		GB_PR4:AC007032	126803	AC007032	Homo sapiens clone NH0022N19, complete sequence.	Homo sapiens	34,180	17-Jul-99
		GB_PR4:AC007032	126803	AC007032	Homo sapiens clone NH0022N19, complete sequence.	Homo sapiens	36,858	17-Jul-99
rxa01428	1260	GB_BA1:SCH24	41625	AL049826	Streptomyces coelicolor cosmid H24.	Streptomyces coelicolor	51,278	11-MAY-1999
		GB_BA2:AF031590	6676	AF031590	Streptomyces coelicolor thioredoxin (trxA) gene, partial cds; SpoOJ-like, Soj-like, GidB-like, Jag-like, inner membrane protein, and 9-10kDa protein-like genes, complete cds; RNase P protein (rnpA) gene, partial cds; and unknown gene.	Streptomyces coelicolor	39,389	20-Feb-98
		GB_BA1:SCTRXARNP	6676	Y16311	Streptomyces coelicolor trxA & rnpA genes & ORFs 205, 344, 255, 239, 170, 341 & 124.	Streptomyces coelicolor	39,389	18-DEC-1998
rxa01430	1311	GB_EST30:AI643302	254	AI643302	vI39b08.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:974583 5' similar to SW:6PGD_HUMAN P52209 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING ;, mRNA sequence.	Mus musculus	38,627	29-Apr-99

TABLE 4: ALIGNMENT RESULTS

		GB_EST34:AI788121	490	AI788121	ul17f02.y1 Sugano mouse embryo mewa Mus musculus cDNA clone IMAGE:2087835 5' similar to SW:6PGD_HUMAN P52209 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING ;, mRNA sequence.	Mus musculus	40,583	2-Jul-99
		GB_EST16:AA560354	253	AA560354	vl39b08.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:974583 5' similar to TR:G984325 G984325 PHOSPHOGLUCONATE DEHYDROGENASE. ;, mRNA sequence.	Mus musculus	42,544	18-Aug-97
rxa01435	893	GB_EST22:AI069195	892	AI069195	mgae0005dF02f Magnaporthe grisea Appressorium Stage cDNA Library Pyricularia grisea cDNA clone mgae0005dF02f 5', mRNA sequence.	Pyricularia grisea	40,964	09-DEC-1999
		GB_EST26:AI392390	574	AI392390	NCSC1B12T7 Subtracted Conidial Neurospora crassa cDNA clone SC1B12 3' similar to adenylate kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.	Neurospora crassa	40,127	3-Feb-99
		GB_HTG2:AC004845	140230	AC004845	Homo sapiens clone DJ0635O05, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	Homo sapiens	36,437	12-Jun-98
rxa01437	1506	GB_BA1:CGPTAACKA	3657	X89084	C.glutamicum pta gene and ackA gene.	Corynebacterium glutamicum	100,000	23-MAR-1999
		GB_BA1:MTCY22G10	35420	Z84724	Mycobacterium tuberculosis H37Rv complete genome; segment 21/162.	Mycobacterium tuberculosis	54,867	17-Jun-98
		GB_HTG3:AC010254	114363	AC010254	Homo sapiens chromosome 5 clone CIT-HSPC_434O11, *** SEQUENCING IN PROGRESS ***, 58 unordered pieces.	Homo sapiens	35,547	15-Sep-99
rxa01461	735	GB_BA2:AF003947	5475	AF003947	Rhodococcus opacus succinyl CoA:3-oxoadipate CoA transferase subunit homolog (pcaI') gene, partial cds, protocatechuate dioxygenase beta subunit (pcaH), protocatechuate dioxygenase alpha subunit (pcaG), 3-carboxy-cis,cis-muconate cycloisomerase homolog (pcaB), 3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase (pcaL) and PcaR (pcaR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcaF') gene, partial cds.	Rhodococcus opacus	57,939	12-MAR-1998
		GB_PR2:HSA535K18	182408	AL078638	Human DNA sequence from clone RP11-535K18 on chromosome Xq26.2-27.1, complete sequence.	Homo sapiens	37,123	22-Nov-99
		GB_EST33:AI764654	420	AI764654	UI-R-Y0-abw-e-02-0-UI.s2 UI-R-Y0 Rattus norvegicus cDNA clone UI-R-Y0-abw-e-02-0-UI 3', mRNA sequence.	Rattus norvegicus	35,885	25-Jun-99
rxa01462	813	GB_BA2:AF003947	5475	AF003947	Rhodococcus opacus succinyl CoA:3-oxoadipate CoA transferase subunit homolog (pcaI') gene, partial cds, protocatechuate dioxygenase beta subunit (pcaH), protocatechuate dioxygenase alpha subunit (pcaG), 3-carboxy-cis,cis-muconate cycloisomerase homolog (pcaB), 3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase (pcaL) and PcaR (pcaR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcaF') gene, partial cds.	Rhodococcus opacus	66,667	12-MAR-1998
		GB_BA1:SC4C6	30941	AL079355	Streptomyces coelicolor cosmid 4C6.	Streptomyces coelicolor	40,822	21-Jun-99
		GB_BA2:AF109386	6551	AF109386	Streptomyces sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	Streptomyces sp. 2065	56,049	06-DEC-1999
rxa01464	414	GB_BA1:AB009343	6342	AB009343	Frateuria sp. ANA-18 ORFR2, catBI, catCI, catAI and catD genes, complete cds.	Frateuria sp. ANA-18	50,966	26-MAY-1999
		GB_GSS10:AQ241375	284	AQ241375	CITBI-E1-2505O7.TF.1 CITBI-E1 Homo sapiens genomic clone 2505O7, genomic survey sequence.	Homo sapiens	39,085	30-Sep-98

TABLE 4: ALIGNMENT RESULTS

		GB_HTG3:AC010363	174962	AC010363	Homo sapiens chromosome 5 clone CITB-H1_2039P12, *** SEQUENCING IN PROGRESS ***, 43 unordered pieces.	Homo sapiens	35,784	15-Sep-99
rxa01465	1284	GB_BA1:ROX99622	7224	X99622	Rhodococcus opacus catR, catA, catB, catC genes and five ORFs.	Rhodococcus opacus	58,814	24-Sep-97
		GB_BA1:D83237	1626	D83237	Rhodococcus erythropolis DNA for catechol 1,2-dioxgenase, complete cds.	Rhodococcus erythropolis	53,904	1-Sep-99
		GB_EST9:AA119571	445	AA119571	mp68d04.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:574375 5' similar to TR:G559375 G559375 RAS GTPASE-ACTIVATING PROTEIN. ;, mRNA sequence.	Mus musculus	39,551	17-Feb-97
rxa01466	1083	GB_EST37:AI934978	425	AI934978	wd17b06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2328371 3', mRNA sequence.	Homo sapiens	43,609	2-Sep-99
		GB_EST15:AA465729	289	AA465729	aa32g06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:815002 3', mRNA sequence.	Homo sapiens	41,115	13-Aug-97
		GB_EST24:AI219091	633	AI219091	qg12a05.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1759280 3' similar to TR:Q99988 Q99988 TGF-BETA SUPERFAMILY PROTEIN. [1] ;, mRNA sequence.	Homo sapiens	36,066	29-Nov-98
rxa01477	1671	GB_BA2:CGU89648	1105	U89648	Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence.	Corynebacterium glutamicum	49,726	30-MAR-1999
		GB_EST21:AA919685	782	AA919685	vx11g06.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:1264186 5' similar to gb:M73696 Murine Glvr-1 mRNA, complete cds (MOUSE);, mRNA sequence.	Mus musculus	37,762	20-Apr-98
		GB_HTG2:HS1005F21	101795	AL078633	Homo sapiens chromosome 20 clone RP5-1005F21, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	38,371	30-Nov-99
rxa01499	3945	GB_PR4:AC006454	153201	AC006454	Homo sapiens clone DJ0852P06, complete sequence.	Homo sapiens	38,033	13-Aug-99
		GB_BA1:LSLYSSYNT	4724	X96558	Lysobacter sp. gene encoding synthetase.	Lysobacter	42,840	8-Jan-97
		GB_PR4:AC006454	153201	AC006454	Homo sapiens clone DJ0852P06, complete sequence.	Homo sapiens	38,823	13-Aug-99
rxa01502	1356	GB_PAT:I92046	2203	I92046	Sequence 13 from patent US 5726299.	Unknown.	39,755	01-DEC-1998
		GB_PAT:I78757	2203	I78757	Sequence 13 from patent US 5693781.	Unknown.	39,755	3-Apr-98
		GB_BA1:MTCY359	36021	Z83859	Mycobacterium tuberculosis H37Rv complete genome; segment 84/162.	Mycobacterium tuberculosis	36,613	17-Jun-98
rxa01509	597	GB_BA1:SCE9	37730	AL049841	Streptomyces coelicolor cosmid E9.	Streptomyces coelicolor	60,637	19-MAY-1999
		GB_BA1:MTY15C10	33050	Z95436	Mycobacterium tuberculosis H37Rv complete genome; segment 154/162.	Mycobacterium tuberculosis	59,296	17-Jun-98
		GB_BA1:MLCB2548	38916	AL023093	Mycobacterium leprae cosmid B2548.	Mycobacterium leprae	59,764	27-Aug-99
rxa01510	1404	GB_GSS9:AQ129927	440	AQ129927	HS_2165_B1_D09_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2165 Col=17 Row=H, genomic survey sequence.	Homo sapiens	36,136	23-Sep-98
		GB_BA2:AF016585	41097	AF016585	Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi) gene, partial cds; polyketide synthase modules 1 through 7 (nidA) genes, complete cds; and N-methyltransferase homolog gene, partial cds.	Streptomyces caelestis	37,464	07-DEC-1997
		GB_HTG4:AC010747	216500	AC010747	Homo sapiens chromosome unknown clone NH0555H09, WORKING DRAFT SEQUENCE, in unordered pieces.	Homo sapiens	33,022	29-OCT-1999
rxa01511	1065	GB_BA1:BRLBIOBA	1647	D14084	Brevibacterium flavum gene for biotin synthetase, complete cds.	Corynebacterium glutamicum	40,283	3-Feb-99
		GB_GSS3:B45213	358	B45213	HS-1060-B2-D07-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 782 Col=14 Row=H, genomic survey sequence.	Homo sapiens	49,505	21-OCT-1997

TABLE 4: ALIGNMENT RESULTS

		GB_HTG4:AC010747	216500	AC010747	Homo sapiens chromosome unknown clone NH0555H09, WORKING DRAFT SEQUENCE, in unordered pieces.	Homo sapiens	33,819	29-OCT-1999
rxa01513	2682	GB_BA1:MTCY7H7B	24244	Z95557	Mycobacterium tuberculosis H37Rv complete genome; segment 153/162.	Mycobacterium tuberculosis	40,354	18-Jun-98
		GB_BA2:AF037269	2364	AF037269	Mycobacterium smegmatis cell division protein (FtsH) gene, complete cds.	Mycobacterium smegmatis	60,814	19-Aug-98
		GB_BA1:MLCB2548	38916	AL023093	Mycobacterium leprae cosmid B2548.	Mycobacterium leprae	39,992	27-Aug-99
rxa01593	990	GB_BA1:U00012	33312	U00012	Mycobacterium leprae cosmid B1308.	Mycobacterium leprae	39,126	30-Jan-96
		GB_IN1:CELF27E11	25700	AF016413	Caenorhabditis elegans cosmid F27E11.	Caenorhabditis elegans	34,227	2-Aug-97
		GB_OV:DYGAGR	4354	L01423	Discopyge ommata (clone OL4) agrin mRNA, 3' end cds.	Discopyge ommata	38,414	28-Apr-93
rxa01608	1962	GB_BA2:AF119150	18605	AF119150	Vibrio cholerae Rtx toxin gene cluster, complete cds.	Vibrio cholerae	36,919	21-MAR-1999
		GB_BA2:AF119150	18605	AF119150	Vibrio cholerae Rtx toxin gene cluster, complete cds.	Vibrio cholerae	38,130	21-MAR-1999
rxa01620								
rxa01640	3441	GB_PR3:HS52D1	148691	Z96811	Human DNA sequence from PAC 52D1 on chromosome Xq21. Contains CA repeats, STS.	Homo sapiens	35,501	23-Nov-99
		GB_BA2:AF079155	686	AF079155	Ralstonia eutropha phasin (phaP) mRNA, complete cds.	Ralstonia eutropha	40,497	6-Apr-99
		GB_IN2:AF039570	1866	AF039570	Caenorhabditis elegans aryl hydrocarbon receptor ortholog AHR-1 (ahr-1) mRNA, complete cds.	Caenorhabditis elegans	39,699	04-OCT-1999
rxa01653	1584	GB_HTG7:AC010997	187768	AC010997	Homo sapiens clone RP11-399K21, *** SEQUENCING IN PROGRESS ***, 35 unordered pieces.	Homo sapiens	34,516	08-DEC-1999
		GB_HTG7:AC010997	187768	AC010997	Homo sapiens clone RP11-399K21, *** SEQUENCING IN PROGRESS ***, 35 unordered pieces.	Homo sapiens	36,177	08-DEC-1999
		GB_VI:AF030154	34446	AF030154	Bovine adenovirus 3 complete genome.	bovine adenovirus type 3	40,345	27-Jan-99
rxa01716	509	GB_BA1:AB010645	16836	AB010645	Acetobacter xylinus genes for endoglucanase, cellulose synthase subunit ABCD and beta-glucosidase, complete cds.	Acetobacter xylinus	34,783	13-Feb-99
		GB_BA1:AB010645	16836	AB010645	Acetobacter xylinus genes for endoglucanase, cellulose synthase subunit ABCD and beta-glucosidase, complete cds.	Acetobacter xylinus	37,598	13-Feb-99
		GB_BA1:ABCBCSABCD	9540	M37202	A.xylinum bcs A, B, C and D genes, complete cds's.	Acetobacter xylinus	39,173	24-Apr-93
rxa01728	1098	GB_BA2:CORCSLYS	2821	M89931	Corynebacterium glutamicum beta C-S lyase (aecD) and branched-chain amino acid uptake carrier (brnQ) genes, complete cds, and hypothetical protein Yhbw (yhbw) gene, partial cds.	Corynebacterium glutamicum	99,636	4-Jun-98
		GB_PL2:HAAP	931	X95952	H.annuus mRNA for aquaporin.	Helianthus annuus	39,231	14-Jul-99
		GB_HTG1:CEY32F6	187816	AL008875	Caenorhabditis elegans chromosome V clone Y32F6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	37,431	9-Nov-97
rxa01732	1173	GB_PR4:HUAC004125	194020	AC004125	Homo sapiens Chromosome 16 BAC clone CIT987SK-625P11, complete sequence.	Homo sapiens	35,345	23-Nov-99
		GB_PR4:HUAC004125	194020	AC004125	Homo sapiens Chromosome 16 BAC clone CIT987SK-625P11, complete sequence.	Homo sapiens	37,381	23-Nov-99
		GB_IN1:CER11A5	26671	Z83122	Caenorhabditis elegans cosmid R11A5, complete sequence.	Caenorhabditis elegans	36,140	2-Sep-99
rxa01810	1200	GB_EST28:AI499508	403	AI499508	to02d01.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2177857 3' similar to SW:NU4M_PANTR P03906 NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 ;, mRNA sequence.	Homo sapiens	36,725	11-MAR-1999

TABLE 4: ALIGNMENT RESULTS

		GB_EST28:AI499508	403	AI499508	to02d01.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2177857 3' similar to SW:NU4M_PANTR P03906 NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 ;, mRNA sequence.	Homo sapiens	38,264	11-MAR-1999
rxa01828	1545	GB_BA1:MLCB1770	37821	Z70722	Mycobacterium leprae cosmid B1770.	Mycobacterium leprae	36,411	29-Aug-97
		GB_HTG2:AC008073	173144	AC008073	Homo sapiens clone NH0507M03, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	36,310	17-Jul-99
		GB_HTG2:AC008073	173144	AC008073	Homo sapiens clone NH0507M03, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	36,310	17-Jul-99
rxa01829	1446	GB_IN1:AB018544	620	AB018544	Hydra magnipapillata mRNA for Hym-176 preprohormone, complete cds.	Hydra magnipapillata	34,855	6-Feb-99
		GB_EST8:AA003136	450	AA003136	mg51e01.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:427320 5' similar to gb:X07315 PLACENTAL PROTEIN 15 (HUMAN);, mRNA sequence.	Mus musculus	42,202	19-Jul-96
		GB_IN1:AB018544	620	AB018544	Hydra magnipapillata mRNA for Hym-176 preprohormone, complete cds.	Hydra magnipapillata	35,968	6-Feb-99
rxa01868	2049	GB_BA1:MTV033	21620	AL021928	Mycobacterium tuberculosis H37Rv complete genome; segment 11/162.	Mycobacterium tuberculosis	38,679	17-Jun-98
		GB_BA1:MLCL622	42498	Z95398	Mycobacterium leprae cosmid L622.	Mycobacterium leprae	38,911	24-Jun-97
		GB_BA1:MSGB983CS	36788	L78828	Mycobacterium leprae cosmid B983 DNA sequence.	Mycobacterium leprae	38,933	15-Jun-96
rxa01934	681	GB_PR4:DJ534K4	216387	AF109907	Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds.	Homo sapiens	39,189	23-DEC-1998
		GB_HTG2:AC006342	201618	AC006342	Homo sapiens clone DJ0054D12, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	34,412	11-Jan-99
		GB_HTG2:AC006342	201618	AC006342	Homo sapiens clone DJ0054D12, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	34,412	11-Jan-99
rxa01967	1266	GB_IN2:AC005467	62091	AC005467	Drosophila melanogaster, chromosome 2R, region 48C1-48C2, P1 clone DS00568, complete sequence.	Drosophila melanogaster	35,252	12-DEC-1998
		GB_BA2:AE001678	13485	AE001678	Chlamydia pneumoniae section 94 of 103 of the complete genome.	Chlamydophila pneumoniae	35,203	08-MAR-1999
		GB_IN2:AC005467	62091	AC005467	Drosophila melanogaster, chromosome 2R, region 48C1-48C2, P1 clone DS00568, complete sequence.	Drosophila melanogaster	34,699	12-DEC-1998
rxa01993	1166	GB_BA1:PPVANAB	2864	Y14759	Pseudomonas putida vanA and vanB genes.	Pseudomonas putida	51,697	09-MAY-1998
		GB_HTG2:AC006799	278007	AC006799	Caenorhabditis elegans clone Y51H7, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	Caenorhabditis elegans	38,455	23-Feb-99
		GB_HTG2:AC006799	278007	AC006799	Caenorhabditis elegans clone Y51H7, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	Caenorhabditis elegans	38,455	23-Feb-99
rxa01994	1098	GB_HTG4:AC009961	231522	AC009961	Homo sapiens chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unordered pieces.	Homo sapiens	35,576	29-OCT-1999
		GB_HTG4:AC009961	231522	AC009961	Homo sapiens chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unordered pieces.	Homo sapiens	35,576	29-OCT-1999
		GB_HTG4:AC009961	231522	AC009961	Homo sapiens chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unordered pieces.	Homo sapiens	35,472	29-OCT-1999
rxa01997	609	GB_BA2:AF112536	1798	AF112536	Corynebacterium glutamicum ribonucleotide reductase beta-chain (nrdF) gene, complete cds.	Corynebacterium glutamicum	37,719	5-Aug-99
		GB_BA1:SCH66	9153	AL049731	Streptomyces coelicolor cosmid H66.	Streptomyces coelicolor	38,655	29-Apr-99
		GB_EST29:AI558691	598	AI558691	fb79c10.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to SW:ATF3_HUMAN P18847 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-3 ;, mRNA sequence.	Danio rerio	40,232	24-MAR-1999

TABLE 4: ALIGNMENT RESULTS

rxa02052	915	GB_EST3:R64206	453	R64206	yi18d08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:139599 5', mRNA sequence.	Homo sapiens	35,920	26-MAY-1995
		GB_PR2:AC002540	70851	AC002540	Human BAC clone GS025M02 from 7q21-q22, complete sequence.	Homo sapiens	37,099	12-Sep-97
		GB_GSS3:B55001	406	B55001	CIT-HSP-385H2.TRB CIT-HSP Homo sapiens genomic clone 385H2, genomic survey sequence.	Homo sapiens	35,599	20-Jun-98
rxa02064	762	GB_PR4:AF135187	33016	AF135187	Homo sapiens interferon-induced protein p78 (MX1) gene, complete cds.	Homo sapiens	32,935	8-Jul-99
		GB_PR3:AC005612	60904	AC005612	Homo sapiens chromosome 21, P1 clone LBL#8 (LBNL H8), complete sequence.	Homo sapiens	32,935	4-Sep-98
		GB_PR1:HUM8DC11Z	3949	L35666	Homo sapiens (subclone H8 10_f11 from P1 35 H5 C8) DNA sequence.	Homo sapiens	31,995	22-Aug-94
rxa02082	3010	GB_BA1:MSGB32CS	36404	L78818	Mycobacterium leprae cosmid B32 DNA sequence.	Mycobacterium leprae	50,604	15-Jun-96
		GB_BA1:MTCY338	29372	Z74697	Mycobacterium tuberculosis H37Rv complete genome; segment 127/162.	Mycobacterium tuberculosis	38,113	17-Jun-98
		GB_GSS10:AQ242118	766	AQ242118	3I23-4r Ochrobactrum anthropi BAC Library Ochrobactrum anthropi genomic clone 3I23-4r, genomic survey sequence.	Ochrobactrum anthropi	41,876	02-OCT-1998
rxa02083	1533	GB_PR4:AC008055	196899	AC008055	Homo sapiens 12q22-103.4-106.5 BAC RPCI11-718L23 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	Homo sapiens	36,818	09-OCT-1999
		GB_PL2:AC002292	120787	AC002292	Genomic sequence of Arabidopsis BAC F8A5, complete sequence.	Arabidopsis thaliana	37,517	02-OCT-1997
		GB_PR4:AC008055	196899	AC008055	Homo sapiens 12q22-103.4-106.5 BAC RPCI11-718L23 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	Homo sapiens	35,563	09-OCT-1999
rxa02092	1761	GB_BA2:AF031929	2675	AF031929	Lactobacillus helveticus cochaperonin GroES and chaperonin GroEL genes, complete cds; and DNA mismatch repair enzyme (hexA) gene, partial cds.	Lactobacillus helveticus	36,149	8-Aug-98
		GB_HTG1:HSDJ34F7	129811	AL049547	Homo sapiens chromosome 6 clone RP1-34F7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,587	23-Nov-99
		GB_PR2:HSU24578	17488	U24578	Human RP1 and complement C4B precursor (C4B) genes, partial cds.	Homo sapiens	36,755	16-MAY-1996
rxa02098	1869	GB_BA1:CAJ10319	5368	AJ010319	Corynebacterium glutamicum amtP, glnB, glnD genes and partial ftsY and srp genes.	Corynebacterium glutamicum	99,766	14-MAY-1999
		GB_BA1:CAJ10319	5368	AJ010319	Corynebacterium glutamicum amtP, glnB, glnD genes and partial ftsY and srp genes.	Corynebacterium glutamicum	36,983	14-MAY-1999
rxa02105	391	GB_EST17:AA660065	352	AA660065	EST00115 watermelon lambda zap express library Citrullus lanatus cDNA clone WMLS233 5' similar to translation initiation factor, mRNA sequence.	Citrullus lanatus	37,231	10-Nov-97
		GB_GSS6:AQ839377	523	AQ839377	HS_4640_B2_F09_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=4640 Col=18 Row=L, genomic survey sequence.	Homo sapiens	37,500	30-Aug-99
		GB_PL1:SPCC970	31438	AL031530	S.pombe chromosome III cosmid c970.	Schizosaccharomyces pombe	38,268	07-MAY-1999
rxa02111	1407	GB_BA1:SC6G10	36734	AL049497	Streptomyces coelicolor cosmid 6G10.	Streptomyces coelicolor	50,791	24-MAR-1999
		GB_BA1:U00010	41171	U00010	Mycobacterium leprae cosmid B1170.	Mycobacterium leprae	37,563	01-MAR-1994
		GB_BA1:MTCY336	32437	Z95586	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.	Mycobacterium tuberculosis	39,504	24-Jun-99
rxa02118	465	GB_HTG2:AC007164	158320	AC007164	Homo sapiens clone NH0304A10, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,377	23-Apr-99
		GB_PR3:AC004451	108642	AC004451	Homo sapiens PAC clone DJ0789N01 from 7q21, complete sequence.	Homo sapiens	39,387	20-MAR-1998
		GB_HTG2:AC007164	158320	AC007164	Homo sapiens clone NH0304A10, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,377	23-Apr-99

TABLE 4: ALIGNMENT RESULTS

rxa02120	885	GB_PL2:PUMCDC2A	1288	L34206	Petroselinum crispum protein kinase p34cdc2 (cdc2) mRNA, complete cds.	Petroselinum crispum	37,816	17-Feb-96
		GB_GSS10:AQ214799	431	AQ214799	HS_3010_A2_G12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3010 Col=24 Row=M, genomic survey sequence.	Homo sapiens	34,591	18-Sep-98
		GB_PL2:PUMCDC2A	1288	L34206	Petroselinum crispum protein kinase p34cdc2 (cdc2) mRNA, complete cds.	Petroselinum crispum	36,541	17-Feb-96
rxa02126	444	GB_GSS4:AQ707596	485	AQ707596	HS_5560_B1_H08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1136 Col=15 Row=P, genomic survey sequence.	Homo sapiens	38,482	7-Jul-99
		GB_GSS13:AQ494885	411	AQ494885	HS_5195_A1_B11_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=771 Col=21 Row=C, genomic survey sequence.	Homo sapiens	40,897	28-Apr-99
		GB_GSS4:AQ707596	485	AQ707596	HS_5560_B1_H08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1136 Col=15 Row=P, genomic survey sequence.	Homo sapiens	43,533	7-Jul-99
rxa02148	1266	GB_HTG2:AC007905	100722	AC007905	Homo sapiens chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	Homo sapiens	36,051	24-Jun-99
		GB_HTG2:AC007905	100722	AC007905	Homo sapiens chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	Homo sapiens	36,051	24-Jun-99
		GB_HTG2:AC007905	100722	AC007905	Homo sapiens chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	Homo sapiens	35,402	24-Jun-99
rxa02214	732	GB_GSS13:AQ459868	402	AQ459868	HS_5116_A1_H04_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=692 Col=7 Row=O, genomic survey sequence.	Homo sapiens	43,035	23-Apr-99
		GB_EST26:AU005050	790	AU005050	AU005050 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ws30188, mRNA sequence.	Bombyx mori	45,902	19-Jan-99
		GB_PL2:F8K7	98581	AC007727	Arabidopsis thaliana chromosome 1 BAC F8K7 sequence, complete sequence.	Arabidopsis thaliana	37,155	29-Jun-99
rxa02316	1137	GB_EST32:AI723424	600	AI723424	hcgls49.T7 Haemonchus contortus Intestinal mRNA Haemonchus contortus cDNA clone hcgls49.T7 T7, mRNA sequence.	Haemonchus contortus	35,953	10-Jun-99
		GB_PR4:AC000134	203300	AC000134	Homo sapiens Chromosome 11q13 BAC Clone 137c7, complete sequence.	Homo sapiens	37,030	06-MAY-1999
		GB_STS:AF021124	575	AF021124	Homo sapiens trinucleotide repeat ctg-68, sequence tagged site.	Homo sapiens	41,913	3-Apr-98
rxa02384	831	GB_PL1:ATA224957	4081	AJ224957	Arabidopsis thaliana RGAL gene.	Arabidopsis thaliana	35,627	19-MAY-1998
		GB_RO:AF022770	577	AF022770	Mus musculus peripherial benzodiazepine receptor associated protein (Pap7) mRNA, partial cds.	Mus musculus	39,652	24-Sep-97
		GB_GSS11:AQ258908	890	AQ258908	nbxb0021F23r CUGI Rice BAC Library Oryza sativa genomic clone nbxb0021F23r, genomic survey sequence.	Oryza sativa	39,515	23-OCT-1998
rxa02411	972	GB_BA1:AB020624	1605	AB020624	Corynebacterium glutamicum murI gene for D-glutamate racemase, complete cds.	Corynebacterium glutamicum	98,868	24-Jul-99
		GB_EST18:AA733776	385	AA733776	vv03f03.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1210589 5', mRNA sequence.	Mus musculus	43,864	7-Jan-98
		GB_EST38:AW033449	612	AW033449	EST277020 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC28F5, mRNA sequence.	Lycopersicon esculentum	35,620	15-Sep-99

TABLE 4: ALIGNMENT RESULTS

rxa02448	1212	GB_BA1:AB016258	2260	AB016258	Arthrobacter sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	Arthrobacter sp.	60,465	8-Sep-99
		GB_EST37:AW014148	553	AW014148	UI-H-BI0-aaj-c-04-0-UI.s1 NCI_CGAP_Sub1 Homo sapiens cDNA clone IMAGE:2709487 3', mRNA sequence.	Homo sapiens	44,560	10-Sep-99
		GB_EST14:AA432042	543	AA432042	zw80f01.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782521 5' similar to WP:T12A7.1 CE06433 ;, mRNA sequence.	Homo sapiens	36,522	22-MAY-1997
rxa02449	1026	GB_BA1:AB016258	2260	AB016258	Arthrobacter sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	Arthrobacter sp.	66,244	8-Sep-99
		GB_BA1:CGPUTP	3791	Y09163	C.glutamicum putP gene.	Corynebacterium glutamicum	39,899	8-Sep-97
		GB_BA1:AB016258	2260	AB016258	Arthrobacter sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	Arthrobacter sp.	70,410	8-Sep-99
rxa02497	1050	GB_BA2:CGU31224	422	U31224	Corynebacterium glutamicum (ppx) gene, partial cds.	Corynebacterium glutamicum	96,445	2-Aug-96
		GB_BA1:MTCY20G9	37218	Z77162	Mycobacterium tuberculosis H37Rv complete genome; segment 25/162.	Mycobacterium tuberculosis	59,429	17-Jun-98
		GB_BA1:SCE7	16911	AL049819	Streptomyces coelicolor cosmid E7.	Streptomyces coelicolor	39,510	10-MAY-1999
rxa02526	1329	GB_GSS10:AQ240233	483	AQ240233	CIT-HSP-2385F9.TR.1 CIT-HSP Homo sapiens genomic clone 2385F9, genomic survey sequence.	Homo sapiens	42,475	30-Sep-98
		GB_OV:S48556	195	S48556	{tandem repeat P1 monomer} [Cacatua galerita=sulfur-crested cockatoo, Genomic, 195 nt].	Cacatua galerita	50,515	08-MAY-1993
		GB_PR2:HSM801056	2555	AL117532	Homo sapiens mRNA; cDNA DKFZp434E192 (from clone DKFZp434E192).	Homo sapiens	39,116	15-Sep-99
rxa02530	780	GB_PR3:HSJ753D10	97912	AL049651	Human DNA sequence from clone 753D10 on chromosome 20 Contains genes for SSTR4(somatostatin receptor 4) and THBD(thrombomodulin), ESTs, STSs, GSSs and CpG islands, complete sequence.	Homo sapiens	34,248	23-Nov-99
		GB_EST33:AI782764	661	AI782764	EST263643 tomato susceptible, Cornell Lycopersicon esculentum cDNA clone cLES20B10, mRNA sequence.	Lycopersicon esculentum	35,385	29-Jun-99
		GB_GSS9:AQ121479	521	AQ121479	HS_3084_A2_B02_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3084 Col=4 Row=C, genomic survey sequence.	Homo sapiens	38,689	22-Sep-98
rxa02535	1278	GB_HTG3:AC008710	146065	AC008710	Homo sapiens chromosome 5 clone CIT978SKB_7E3, *** SEQUENCING IN PROGRESS ***, 39 unordered pieces.	Homo sapiens	35,799	3-Aug-99
		GB_HTG3:AC008710	146065	AC008710	Homo sapiens chromosome 5 clone CIT978SKB_7E3, *** SEQUENCING IN PROGRESS ***, 39 unordered pieces.	Homo sapiens	35,799	3-Aug-99
		GB_HTG3:AC008710	146065	AC008710	Homo sapiens chromosome 5 clone CIT978SKB_7E3, *** SEQUENCING IN PROGRESS ***, 39 unordered pieces.	Homo sapiens	34,886	3-Aug-99
rxa02603	1119	GB_BA1:MTV026	23740	AL022076	Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.	Mycobacterium tuberculosis	37,975	24-Jun-99
		GB_IN2:AC005714	177740	AC005714	Drosophila melanogaster, chromosome 2R, region 58D4-58E2, BAC clone BACR48M13, complete sequence.	Drosophila melanogaster	41,226	01-MAY-1999
		GB_EST19:AA775050	218	AA775050	ac76e10.s1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:868554 3' similar to gb:Y00371_rna1 HEAT SHOCK COGNATE 71 KD PROTEIN (HUMAN);, mRNA sequence.	Homo sapiens	40,826	5-Feb-98

rxa02641

TABLE 4: ALIGNMENT RESULTS

rxa02651	1053	GB_BA1:MTCY48	35377	Z74020	Mycobacterium tuberculosis H37Rv complete genome; segment 69/162.	Mycobacterium tuberculosis	62,678	17-Jun-98
		GB_BA1:SC4A10	43147	AL109663	Streptomyces coelicolor cosmid 4A10.	Streptomyces coelicolor A3(2)	39,109	5-Aug-99
		GB_BA1:MLCL458	43839	AL049478	Mycobacterium leprae cosmid L458.	Mycobacterium leprae	62,753	27-Aug-99
rxa02674	1575	GB_BA2:PPU96338	5276	U96338	Pseudomonas putida NCIMB 9866 plasmid pRA4000 p-cresol degradative pathway genes, p-hydroxybenzaldehyde dehydrogenase (pchA), p-cresol methylhydroxylase, cytochrome subunit precursor (pchC), unknown (pchX) and p-cresol methylhydroxylase, flavoprotein subunit (pchF) genes, complete cds.	Pseudomonas putida	58,095	13-MAY-1999
		GB_BA1:SCE9	37730	AL049841	Streptomyces coelicolor cosmid E9.	Streptomyces coelicolor	38,544	19-MAY-1999
		GB_BA2:PPU96339	4464	U96339	Pseudomonas putida NCIMB 9869 plasmid pRA500 p-cresol degradative pathway genes, p-hydroxybenzaldehyde dehydrogenase (pchA) gene, partial cds, and p-cresol methylhydroxylase, cytochrome subunit (pchC), unknown (pchX), p-cresol methylhydroxylase, flavoprotein subunit (pchF), protocatechuate-3,4-dioxygenase, beta subunit (pcaH) and protocatechuate-3,4-dioxygenase, alpha subunit (pcaG) genes, complete cds.	Pseudomonas putida	70,588	13-MAY-1999
rxa02702	1581	GB_BA1:AB015023	2291	AB015023	Corynebacterium glutamicum genes for MurC and FtsQ, complete cds.	Corynebacterium glutamicum	99,365	6-Feb-99
		GB_BA1:AB003132	4116	AB003132	Corynebacterium glutamicum gene for MurC, FtsQ, FtsZ, complete cds.	Corynebacterium glutamicum	99,317	4-Aug-97
		GB_BA1:BLFTSZ	5546	Y08964	B.lactofermentum murC, ftsQ or divD & ftsZ genes.	Corynebacterium glutamicum	99,296	08-OCT-1998
rxa02703	1212	GB_BA1:AB015023	2291	AB015023	Corynebacterium glutamicum genes for MurC and FtsQ, complete cds.	Corynebacterium glutamicum	97,468	6-Feb-99
		GB_PL2:VFAMACTRA	1879	Y09591	V.faba mRNA for amino acid transporter.	Vicia faba	38,915	02-DEC-1999
		GB_PAT:E05047	966	E05047	DNA encoding recombinant monoglyceride lipase.	Bacillus sp.	37,158	29-Sep-97
rxa02704	1812	GB_BA1:MTCY270	37586	Z95388	Mycobacterium tuberculosis H37Rv complete genome; segment 96/162.	Mycobacterium tuberculosis	37,946	10-Feb-99
		GB_BA2:AE000961	18765	AE000961	Archaeoglobus fulgidus section 146 of 172 of the complete genome.	Archaeoglobus fulgidus	38,521	15-DEC-1997
		GB_BA1:MTCY270	37586	Z95388	Mycobacterium tuberculosis H37Rv complete genome; segment 96/162.	Mycobacterium tuberculosis	37,850	10-Feb-99
rxa02705	1539	GB_PAT:I26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	97,619	07-OCT-1996
		EM_PAT:E11760	6911	E11760	Base sequence of sucrase gene.	Corynebacterium glutamicum	97,619	08-OCT-1997 (Rel. 52, Created)
		GB_BA1:SC4A10	43147	AL109663	Streptomyces coelicolor cosmid 4A10.	Streptomyces coelicolor A3(2)	37,856	5-Aug-99
rxa02706	1221	GB_PAT:I26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	98,605	07-OCT-1996
		EM_PAT:E11760	6911	E11760	Base sequence of sucrase gene.	Corynebacterium glutamicum	98,605	08-OCT-1997 (Rel. 52, Created)
		GB_BA1:MTCY270	37586	Z95388	Mycobacterium tuberculosis H37Rv complete genome; segment 96/162.	Mycobacterium tuberculosis	34,868	10-Feb-99
rxa02707	1653	EM_PAT:E11760	6911	E11760	Base sequence of sucrase gene.	Corynebacterium glutamicum	98,547	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:I26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	98,547	07-OCT-1996

TABLE 4: ALIGNMENT RESULTS

		GB_BA1:MLCB268	38859	AL022602	Mycobacterium leprae cosmid B268.	Mycobacterium leprae	37,815	27-Aug-99
rxa02710	1686	EM_PAT:E11760	6911	E11760	Base sequence of sucrase gene.	Corynebacterium glutamicum	52,124	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:I26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	52,124	07-OCT-1996
		GB_GSS13:AQ484169	515	AQ484169	RPCI-11-264A12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-264A12, genomic survey sequence.	Homo sapiens	40,856	24-Apr-99
rxa02711	2235	GB_BA2:XCU45994	1203	U45994	Xanthomonas campestris pv. campestris insertion sequence IS1404.	Xanthomonas campestris pv. campestris	39,061	29-Jan-99
		GB_BA2:XCU77781	4160	U77781	Xanthomonas campestris pv. amaranthicola XamI DNA methyltransferase (xamIM) gene, complete cds; insertion sequence IS1389 and unknown genes.	Xanthomonas campestris pv. amaranthicola	39,551	9-Feb-99
		GB_BA2:AF108355	1222	AF108355	Xanthomonas campestris pv. amaranthicola insertion sequence IS1389-B unknown genes.	Xanthomonas campestris pv. amaranthicola	40,281	09-MAR-1999
rxa02713	1134	GB_BA1:MTCY270	37586	Z95388	Mycobacterium tuberculosis H37Rv complete genome; segment 96/162.	Mycobacterium tuberculosis	38,669	10-Feb-99
		GB_PR1:D31907	599	D31907	Homo sapiens gene for zinc regulatory factor, partial cds.	Homo sapiens	36,396	7-Feb-99
		GB_PR1:HSMTFMR	3302	X78710	H.sapiens MTF-1 mRNA for metal-regulatory transcription factor.	Homo sapiens	37,243	1-Aug-94
rxa02716	684	GB_PR3:AC002347	134977	AC002347	Homo sapiens chromosome 17, clone 297N7, complete sequence.	Homo sapiens	36,282	3-Feb-98
		GB_PR3:HS310J6	87942	AL035593	Human DNA sequence from clone 310J6 on chromosome 6q22.1-22.3. Contains part of a novel gene, ESTs, STSs and GSSs, complete sequence.	Homo sapiens	37,291	23-Nov-99
		GB_HTG3:AC011509	111353	AC011509	Homo sapiens chromosome 19 clone CITB-H1_2189E23, *** SEQUENCING IN PROGRESS ***, 35 unordered pieces.	Homo sapiens	37,407	07-OCT-1999
rxa02722	1449	GB_BA1:BLFTSZ	5546	Y08964	B.lactofermentum murC, ftsQ or divD & ftsZ genes.	Corynebacterium glutamicum	99,652	08-OCT-1998
		GB_BA1:AB003132	4116	AB003132	Corynebacterium glutamicum gene for MurC, FtsQ, FtsZ, complete cds.	Corynebacterium glutamicum	98,535	4-Aug-97
		GB_PAT:E17182	1125	E17182	Brevibacterium flavum ftsQ gene complete cds.	Corynebacterium glutamicum	97,235	28-Jul-99
rxa02723	789	GB_BA1:AB015023	2291	AB015023	Corynebacterium glutamicum genes for MurC and FtsQ, complete cds.	Corynebacterium glutamicum	99,113	6-Feb-99
		GB_BA1:BLFTSZ	5546	Y08964	B.lactofermentum murC, ftsQ or divD & ftsZ genes.	Corynebacterium glutamicum	99,113	08-OCT-1998
		GB_BA1:AB003132	4116	AB003132	Corynebacterium glutamicum gene for MurC, FtsQ, FtsZ, complete cds.	Corynebacterium glutamicum	99,113	4-Aug-97
rxa02813	1108	GB_HTG3:AC009658	171795	AC009658	Homo sapiens chromosome 15 clone 344_A_16 map 15, *** SEQUENCING IN PROGRESS ***, 29 unordered pieces.	Homo sapiens	34,622	01-OCT-1999
		GB_HTG3:AC009658	171795	AC009658	Homo sapiens chromosome 15 clone 344_A_16 map 15, *** SEQUENCING IN PROGRESS ***, 29 unordered pieces.	Homo sapiens	34,622	01-OCT-1999
		GB_RO:MMU65079	2300	U65079	Mus musculus actin-binding protein (ENC-1) mRNA, complete cds.	Mus musculus	35,013	29-Jul-97
rxa02820	1411	GB_BA1:BFU64514	3837	U64514	Bacillus firmus dppABC operon; dipeptide transporter protein dppA gene, partial cds, and dipeptide transporter proteins dppB and dppC genes, complete cds.	Bacillus firmus	36,859	1-Feb-97
		GB_IN1:CET04C10	20958	Z69885	Caenorhabditis elegans cosmid T04C10, complete sequence.	Caenorhabditis elegans	35,934	2-Sep-99

TABLE 4: ALIGNMENT RESULTS

		GB_EST35:AI823090	720	AI823090	L30-944T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-944 5' similar to 60S ribosomal protein L36 (AC004684)[Arabidopsis thaliana], mRNA sequence.	Mesembryanthemum crystallinum	35,770	21-Jul-99
rxa02828	572	GB_BA1:MTCY10H4	39160	Z80233	Mycobacterium tuberculosis H37Rv complete genome; segment 2/162.	Mycobacterium tuberculosis	39,823	17-Jun-98
		GB_BA1:MTORIREP	8400	X92504	M.tuberculosis origin of replication and genes rnpA, rpmH, dnaA, dnaN, recF.	Mycobacterium tuberculosis	39,823	26-Aug-97
		GB_RO:RATENDOGLY	3906	L37380	Rat apical endosomal glycoprotein mRNA, complete cds.	Rattus norvegicus	38,704	20-Apr-95
rxa02839	470	GB_BA2:ECOUW89	176195	U00006	E. coli chromosomal region from 89.2 to 92.8 minutes.	Escherichia coli	99,362	17-DEC-1993
		GB_BA2:AE000477	11314	AE000477	Escherichia coli K-12 MG1655 section 367 of 400 of the complete genome.	Escherichia coli	99,787	12-Nov-98
rxs03218		GB_BA1:ECOPLSB	3865	K00127	E.coli plsB and dgk genes coding for sn-glycerol-3-phosphate acyltransferase and diglyceride kinase.	Escherichia coli	33,761	28-Feb-94